The EY Global Financial Services Institute brings together world-renowned thought leaders and practitioners from top-tier academic institutions, global financial services firms, public policy organizations and regulators to develop solutions to the most pertinent issues facing the financial services industry.

The Journal of Financial Perspectives aims to become the medium of choice for senior financial services executives from banking and capital markets, asset management and insurance, as well as academics and policymakers who wish to keep abreast of the latest ideas from some of the world’s foremost thought leaders in financial services. To achieve this objective, a board comprising leading academic scholars and respected financial executives has been established to solicit articles that not only make genuine contributions to the most important topics, but are also practical in their focus. The Journal will be published three times a year.

gfsi.ey.com
Part 1: Strategic

13 On being the right size
Andrew G Haldane, Executive Director, Financial Stability and Member of the Financial Policy Committee, Bank of England

Commentary: Is big dangerous?
Philip Booth, Editorial and Program Director, Institute of Economic Affairs and Professor of Insurance and Risk Management, Cass Business School, City University

27 Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives
Paula Jarzabkowski, Professor of Strategy, Cass Business School, City University
Rebecca Bednarek, Research Fellow, Cass Business School, City University
Laure Cabantous, Senior Lecturer, Cass Business School, City University

39 In search of a more stable monetary and financial order
Thomas Mayer, Senior Advisor, Deutsche Bank, and Senior Fellow, Center of Financial Studies, Goethe University Frankfurt am Main

47 Rethinking the future of financial services: a structural and evolutionary perspective on regulation
Michael G. Jacobides, Sir Donald Gordon Chair of Entrepreneurship and Innovation and Associate Professor of Strategy and Entrepreneurship, London Business School
Michael Drexler, Senior Director and Head of Investor Industries, World Economic Forum
Jason R. Rico, Senior Project Manager, World Economic Forum

73 E.U. Financial Transaction Tax: a concern for all members of the C-suite
David Williams, Partner, Banking and Capital Markets Advisory, EY LLP
Mark Persoff, Partner, Banking and Capital Markets Tax, EY LLP
Part 2: Tactical

85 Measuring capital adequacy: supervisory stress-tests in a Basel world
Larry D. Wall, Executive Director, Center for Financial Innovation and Stability (CenFIS), Research Department, Federal Reserve Bank of Atlanta

95 Externalities and macroprudential policy
Gianni De Nicolò, Senior Economist, Research Department, International Monetary Fund and Fellow, CESifo
Giovanni Favara, Economist, Federal Reserve Board
Lev Ratnovski, Economist, Research Department, International Monetary Fund

107 Directors’ and officers’ insurance and shareholder protection
M. Martin Boyer, CEFA Professor of Finance and CIRANO Fellow, HEC Montréal, Université de Montréal

129 A long-term approach to Italian banks’ profitability: paradise lost?
Giuseppe Lusignani, Professor of Economics and Financial Intermediaries, University of Bologna
Marco Onado, Senior Professor of Finance, Bocconi University

155 The introduction of market-based pricing in corporate lending
Ivan Ivanov, Financial Economist, Office of Financial Intermediaries, Division of Economic and Risk Analysis, United States Securities and Exchange Commission
João A. C. Santos, Vice President and Function Head, Financial Intermediation Function, Federal Reserve Bank of New York and Full Professor, Nova School of Business and Economics
Thu Vo, Ph.D. Candidate in Finance, Simon School of Business, University of Rochester

161 Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?
Matthias Köhler, Department of Financial Stability, Deutsche Bundesbank

177 Valuation effects of termination of cross-listings
Roland Füss, Full Professor of Real Estate Finance, University of St.Gallen
Ulrich Hommel, Full Professor of Finance, Department of Finance, Accounting and Real Estate, EBS Business School
Jan-Carl Plagge, Ph.D. Candidate, Department of Finance, Accounting and Real Estate, EBS Business School
On being the right size
by Andrew G Haldane, Executive Director, Financial Stability and Member of the Financial Policy Committee, Bank of England

Commentary: Is big dangerous?
by Philip Booth, Editorial and Program Director, Institute of Economic Affairs and Professor of Insurance and Risk Management, Cass Business School, City University

This paper discusses the too-big-to-fail problem from the perspective of individual banks and the financial system as a whole. It starts by exploring the potential implications of recent financial deepening and concentration, which has generated escalating expectations of state support, thereby encouraging further expansion and concentration. The paper then explores three policy approaches to tackling the too-big-to-fail problem. The first is the imposition of systemic surcharges of additional capital, which have the effect of reducing expected system-wide losses in systemically important banks, but not materially so at current levels of the surcharge. Second, new resolution regimes are being put in place to allow banks to fail safely — though the market still has doubts about their credibility for the biggest banks. Finally, structural reform of banks is taking place, through proposals by Volcker, Vickers and Liikanen. Despite this policy progress, expectations of state support remain high. This paper proposes potential additional reforms to tackle too-big-to-fail, such as placing limits on bank size and market share and increasing competition.

Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives
by Paula Jarzabkowski, Professor of Strategy, Cass Business School, City University, Rebecca Bednarek, Research Fellow, Cass Business School, City University and Laure Cabantous, Senior Lecturer, Cass Business School, City University

This paper explores how reinsurers can meet the rapid changes occurring in their industry, arising from primary industry consolidation, and changes in cedent (insurance firm) buyer behavior toward bundled reinsurance products and alternative sources of capital. The paper makes the following suggestions for reinsurers. Reinsurers need to be proactive in responding to changing patterns of premium volume and develop partnerships with global clients. Smaller reinsurers, in particular, will need to look to develop competitive niches and joint-ventures in order to be significant to these large cedents. Furthermore, reinsurers need to continue investing in analytical expertise and resources in order to address the complex needs of their clients. Finally, reinsurers will be increasingly required to engage in alternative risk transfer products, and there will be early-mover advantages in doing so meaningfully.
In search of a more stable monetary and financial order
by Thomas Mayer, Senior Advisor, Deutsche Bank and Senior Fellow, Center of Financial Studies, Goethe University Frankfurt am Main

A popular view is that the “light” regulation of the financial industry during the 1990s and early 2000s allowed the financial excesses and the rise in indebtedness that eventually caused the financial crisis. Those who hold this view want regulation to be tightened to prevent another crisis in the future. In 2008, the G20 demanded that no actor, no product and no market remain unregulated. Since then authorities have been active at the national and international level to fulfill this promise. This paper argues that the premise for more regulation is wrong and hence the suggested therapy misguided. It was not too little but too much official intervention in financial markets that caused the crisis. Instead of more public meddling in private markets we need a stable framework that allows these markets to function properly. In a new framework for monetary policy, credit must be given a prominent role; and in a new framework for regulation, we must properly define safe and loss-absorbing assets and use the latter to cover the costs of bank failures (instead of passing the bill to the tax payers). When the costs of failure are internalized, banks will be under pressure to become more transparent for creditors and depositors and their cost of funds and lending rates will rise. However, the public subsidization of bank lending rates will end only when banks can rely on taxpayer funded bailouts and their efforts to expand so as to become “too big and interconnected to fail” are thwarted.

Rethinking the future of financial services: a structural and evolutionary perspective on regulation
by Michael G. Jacobides, Sir Donald Gordon Chair of Entrepreneurship and Innovation and Associate Professor of Strategy and Entrepreneurship, London Business School, Michael Drexler, Senior Director and Head of Investor Industries, World Economic Forum and Jason R. Rico, Senior Project Manager, World Economic Forum

While the structure of the financial system has undergone a dramatic transformation – to a complex, interconnected, immediate, global sector mediated by capital markets – the fundamental reasons why financial crises occur have changed little over time. This article explains why this transformation went largely unnoticed by academics and regulators, and argues that an outdated view of the financial system that looks at “one piece at a time” and ignores the institutional context may do more harm than good. It articulates an alternative view, showing how a systemic analysis, focused on the structure of the sector and its evolution, can lead to a fresh set of proposals. The article presents our evolutionary structural view with an account of the 2008 financial crisis and preceding events, and assesses the fault-lines that currently remain in our financial system. It concludes with a set of concrete proposals that could improve the stability and performance of the financial sector, and help to reframe both analysis and regulation.
E.U. Financial Transaction Tax: a concern for all members of the C-suite
by David Williams, Partner - Banking and Capital Markets Advisory, EY LLP and
Mark Persoff, Partner - Banking and Capital Markets Tax, EY LLP
In the last two years, French and Italian governments have introduced new taxes on
financial transactions. A proposal for a broader E.U. Financial Transaction Tax exists, and
its design could bring global financial institutions into its reach. The design of the tax
will evolve during 2014; however, even if pared back, the consequences could still have
major financial and operational ramifications for the industry, and will merit discussions
at the Board level. This article explains the origins of the tax, the features of the tax as
currently proposed and what might happen next. It also highlights some of the specific
concerns for members of the Board, including the CEO, CFO, CRO and COO. We conclude
with some practical advice as to what organizations can be doing now to prepare,
recognizing that the scope and timeline for E.U. FTT remains somewhat uncertain.

Measuring capital adequacy: supervisory stress-tests in a Basel world
by Larry D. Wall, Director, Center for Financial Innovation and Stability (CenFIS),
Research Department, Federal Reserve Bank of Atlanta
The U.S. is now committed to using two relatively sophisticated approaches to measuring
capital adequacy: Basel III and stress-tests. The purpose of this paper is to consider
how the two methodologies could work together to help us achieve our supervisory
capital adequacy goals, focusing in particular on how stress-testing could mitigate
weaknesses in Basel III. The paper begins by explaining that the two methodologies take
fundamentally different approaches to measuring a bank’s risk exposure. The Basel III
methodology relies on risk-weighting assets using risk estimates obtained from recent
historical data. The stress-test methodology focuses on the evolution of bank capital
levels in response to a stressful economic scenario. The paper then discusses specific
weaknesses of the Basel III methodology with respect to how it measures credit and
interest rate risk and bank capital, and the way it creates countercyclical capital buffers.
As a part of this discussion, the paper explains how differences in methodologies
potentially allow the stress-tests to mitigate these weaknesses in Basel III. In most cases,
the key difference is that stress-tests give supervisors an added degree of freedom in the
form of being able to specify a stress scenario that addresses a limitation in the Basel
III methodology. However, the paper also emphasizes the extent to which stress-tests
add value will depend upon the exercise of supervisor discretion in the design of stress
scenarios.
Externalities and macroprudential policy
by Gianni De Nicolò, Senior Economist, Research Department, International Monetary Fund and Fellow, CESifo, Giovanni Favara, Economist, Federal Reserve Board and Lev Ratnovski, Economist, Research Department, International Monetary Fund

The recent financial crisis has led to a re-examination of policies applied for achieving macroeconomic and financial stability. Part of the debate involves the adoption of a macroprudential perspective on financial regulation, with the aim of mitigating boom-bust patterns and systemic risks in financial markets. The fundamental rationale for macroprudential policies, however, is not always clearly articulated. The contribution of this paper is to lay out the key sources of market failures that justify macroprudential regulation. It explains how externalities associated with the activity of financial intermediaries can lead to systemic risk, and require specific policies to mitigate such risk. The paper classifies externalities that lead to systemic risk and suggests that the correction of these externalities can be seen as intermediate targets for macroprudential policy, since policies that control externalities mitigate market failures that create systemic risk. This paper discusses how the main proposed macroprudential policy tools—capital requirements, liquidity requirements, restrictions on activities and taxes—address the identified externalities. This mapping may help policymakers select the most effective instruments for macroprudential policy.

Directors’ and officers’ insurance and shareholder protection
by M. Martin Boyer, CEFA Professor of Finance and CIRANO Fellow, HEC Montréal, Université de Montréal

Corporate directors are liable for the corporation’s actions as well as their own. Strangely, and by far, the most likely plaintiffs in a lawsuit against corporate directors are the shareholders who appointed them in the first place. As a result, directors often require protection so that their personal wealth is not expropriated in the event of a good faith error. There are three ways to protect a director’s wealth: corporate indemnification plans, limited liability provision, and directors’ and officers’ (D&O) insurance policies. Of the three types of protection, D&O insurance is arguably the strangest, not because shareholders purchase it to protect directors in case of a lawsuit, but because it also protects shareholders. Despite the extremely popular nature of D&O insurance to protect board members in the U.K., the U.S. and Canada, the study of the drivers of demand for D&O insurance remains a largely unexplored territory as few papers have been devoted to this aspect of corporate governance. Is D&O insurance part of the compensation package? Is it a tool to align the manager’s incentives with those of the shareholders? Is it truly designed to protect corporate directors or other stakeholders, such as the shareholders? Using an original database of Canadian public corporations, I test a set of hypotheses that should determine the demand for D&O insurance. The analysis presented in this article suggests that D&O insurance protects the shareholders’ wealth more than the directors’.
A long-term approach to Italian banks’ profitability: paradise lost?

by Giuseppe Lusignani, Professor of Economics and Financial Intermediaries, University of Bologna and Marco Onado, Senior Professor of Finance, Bocconi University

The long-term profitability of banks is one of the most important issues emerging from the financial crisis. It is now clear that the stellar returns of the last decade (particularly to shareholders, i.e., the return on equity) were the result of decreasing capital and hidden risks (a “mirage”, more than a “miracle” as it has been said). The effect on the banks has been strongly dependent on the business model, the main differences being between investment banks, wholesale banks and retail banks. The problem for retail banks is what will be the profitability of the traditional intermediation business in the foreseeable future. To answer this question, one has to look at the main drivers of profitability in the long term. The Italian banking system is, therefore, an interesting example because it is traditionally retail orientated and underwent in the last two decades a significant process of privatization, consolidation and modernization; so much so that, it proved to be very robust when the financial crisis burst. Moreover, it is one of the few banking systems for which data is available since the late 1960s. The paper shows that today’s low profitability is not only the consequence of the financial crisis, but of a steady decline that began at the end of the 1980s. As the level of profitability proves to be determined mainly by the level of interest rates, a reversal of this trend does not seem likely in the foreseeable future. Consequently, future rebounds of profitability can only come from significant rationalizations, which apparently have not been delivered in the past merger wave, which were meant to achieve economies of scale and scope. This means not only cutting staff costs, but also a radical restructuring of distribution networks and the ways in which customers access bank services.

The introduction of market-based pricing in corporate lending

by Ivan Ivanov, Financial Economist, Office of Financial Intermediaries, Division of Economic and Risk Analysis, United States Securities and Exchange Commission, João A. C. Santos, Vice President and Function Head, Financial Intermediation Division, Federal Reserve Bank of New York and Full Professor, Nova School of Business and Economics and Thu Vo, Ph.D. Candidate in Finance, Simon School of Business, University of Rochester

Prior to the development of the credit default swap (CDS) market, bank loans typically carried a fixed interest rate spread over a base rate, such as prime or LIBOR. The use of fixed-rate spreads, however, requires the inclusion of complex financial covenants and performance-pricing features because borrower credit quality is likely to change during the life of the contract. Over the past 10 years, CDS markets have evolved to reflect borrower-specific credit quality in a timely fashion, giving banks an opportunity to use information contained in CDS spreads in their lending business. Starting in the second quarter of 2008, banks have increasingly extended loans with interest rates tied to borrowers’ CDS spreads or to a CDS index. Drawing on the results in Ivanov, Santos and Vo (2013), this paper discusses the potential outcomes and consequences of this recent financial innovation. Overall, the use of CDS-based interest rate spreads has lowered the cost of bank credit, and it has made bank monitoring less important. However, this innovation could potentially have several adverse consequences, such as reducing the benefits of relationship lending and creating spirals in borrowers’ CDS spreads and loan interest rates.
Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

by Matthias Köhler, Department of Financial Stability, Deutsche Bundesbank

This article provides a brief overview of how business models developed in banking and outlines possible strategies with which banks can restructure their business models in the post-crisis period. At present, it is particularly large, investment-oriented banks that are under pressure to change their business models, since many business areas that guaranteed high returns pre-crisis have become unprofitable. To ensure a sufficient level of profitability, these banks have to refocus on the core competencies and core markets in which they have comparative advantages. While maintaining their focus on corporate and investment banking, they will also need to be more active in retail banking to better diversify their income structure and to meet regulatory requirements. Small, retail-oriented banks have to rethink their business models as well, even though they were much less affected by the 2007/2008 financial crisis. This article shows that these banks will be better off if they increase their share of non-interest income by cross-selling additional services or products to their customers. Diversifying into non-interest income is also beneficial, because competition in retail banking will continue to be high for structural reasons even if interest rates start to go back up. Overall, this paper implies that better-diversified banks are able to generate higher risk-adjusted returns and are more stable. However, while this may reduce idiosyncratic risk, it may also increase the level of systemic risk in the banking sector if banks diversify their activities in a similar way. This suggests that, from a financial stability perspective, it is necessary to have both diversified as well as specialized banks.

Valuation effects of termination of cross-listings

by Roland Füss, Full Professor of Real Estate Finance, University of St. Gallen, Ulrich Hommel, Full Professor of Finance, Department of Finance, Accounting and Real Estate, EBS Business School and Jan-Carl Plagge, Ph.D. Candidate, Department of Finance, Accounting and Real Estate, EBS Business School

In the past, segmentation of capital markets, in particular of developing countries, incentivized numerous companies to cross-list in more integrated and mature markets by promising better investor recognition and thereby lowering cost of capital and increasing stock liquidity. In recent years, however, more and more companies have decided to terminate these dual listings in a foreign country. This paper investigates share price effects linked to the termination of cross-listings. There is a special focus on the characteristics of companies' home markets to explain observed price effects. The empirical results suggest that markets react negatively to the announcement of the termination of the secondary listing, both on the day of the announcement and for a few weeks afterwards. This evaluation even holds for companies that are locally listed on already well-integrated stock exchanges.
Part 1: Strategic

On being the right size

Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives

In search of a more stable monetary and financial order

Rethinking the future of financial services: a structural and evolutionary perspective on regulation

E.U. Financial Transaction Tax: a concern for all members of the C-suite
Abstract
This paper discusses the too-big-to-fail problem from the perspective of individual banks and the financial system as a whole. It starts by exploring the potential implications of recent financial deepening and concentration, which has generated escalating expectations of state support, thereby encouraging further expansion and concentration. The paper then explores three policy approaches to tackling the too-big-to-fail problem. The first is the imposition of systemic surcharges of additional capital, which have the effect of reducing expected system-wide losses in systemically important banks, but not materially so at current levels of the surcharge. Second, new resolution regimes are being put in place to allow banks to fail safely — though the market still has doubts about their credibility for the biggest banks. Finally, structural reform of banks is taking place, through proposals by Volcker, Vickers and Liikanen. Despite this policy progress, expectations of state support remain high. This paper proposes potential additional reforms to tackle too-big-to-fail, such as placing limits on bank size and market share and increasing competition. While existing initiatives are a step in the right direction, there may be some distance to travel before banking is the right size.

1 This article is based on a speech delivered at Institute of Economic Affairs’ 22nd Annual Series, The 2012 Beesley Lectures, at the Institute of Directors, Pall Mall, 25 October 2012. The views are not necessarily those of the Bank of England or the Financial Policy Committee. I would like to thank Andrew Gracie, Varun Paul, Kirsty Rodwell, Timothy Richards, Vicky Saporta, Rhiannon Sowerbutts and Belinda Tracey for their comments and contributions.
1. Introduction

In 1928, evolutionary biologist J. B. S. Haldane wrote an important article whose title I have borrowed, *On Being the Right Size*. The essential point was simple. The sheer size of an object, institution or animal determined their structure. In particular, as their size rose, their structure needed to strengthen more than proportionately if they were to remain robust and resilient. This principle is sometimes enshrined in the so-called “square-cubed” law. A proportional rise in an object’s size causes its area to rise by the square, and its volume by the cube, of that rise. At one level, this is simple mathematical geometry. Yet in the real world, it carries fundamental implications for evolutionary structure.

Take the animal kingdom. The square-cubed law explains why a flea, even if it were the size of a man, would not be capable of jumping to the moon. It explains why a hippopotamus cannot turn somersaults. And it explains why King Kong and Godzilla were physiological impossibilities — the weight transfer associated with a single step would have shattered their thigh bones.

When the world’s biggest banking beasts took a step too far in 2008, they too folded under their own weight. Their physiological structure proved inadequate to make them robust and resilient. That is the essence of the “too-big-to-fail” problem. In the language of Haldane, international policymakers have concluded that many of the world’s largest banks are not the right size given their existing physiological make-up.

Over the past few years, initiatives to solve the too-big-to-fail problem have come thick and fast. At root, each has aimed to strengthen the structure of the world’s biggest banks. That is the good news. Claims that they have solved the too-big-to-fail problem appear to me, however, premature, probably over-optimistic. Worse, they risk sending a false sense of crisis comfort. That is the bad news.

To see why such a cautious conclusion is warranted, we begin by tracking the structural evolution of the financial system over the past few decades. We then consider the three most prominent policy initiatives aimed at tackling too-big-to-fail — systemic surcharges, resolution regimes and structural reform. In the final section, we consider what supplementary policy options might be necessary to ensure banking is right-sized.

2. Evolution of the financial system

The past 50 years have seen seismic shifts in the structure, size and composition of the global financial system. These changes gave birth to the too-big-to-fail problem. Figure 1 plots the ratio of banking sector assets-to-GDP, and its cross-country dispersion, for a set of 14 advanced countries over the past 140 years.

![Figure 1: Bank assets/GDP in selected countries](source: Jorda et al. (2011))

The past 50 years have seen seismic shifts in the structure, size and composition of the global financial system. These changes gave birth to the too-big-to-fail problem. Figure 1 plots the ratio of banking sector assets-to-GDP, and its cross-country dispersion, for a set of 14 advanced countries over the past 140 years.

For the better part of a century, between 1870 and 1970, financial deepening in these countries followed a modestly upward trend. Over this period, the average bank assets-to-GDP ratio rose from 16% to over 70%, or less than six percentage points per decade. Since 1970, this trend has changed trajectory. The ratio of bank assets-to-GDP has more than doubled over the past 40 years, rising from around 70% to over 200%, or over 30 percentage points per decade. In other words, since 1970 financial deepening has occurred five times faster than in the preceding century. For some individual countries, the rise has been more dramatic still — in the U.K., the ratio has risen five-fold.

---

2 Haldane (1928). Haldane is, regrettably, no relative.

3 The data is from Jorda et al. (2011).
In cross-country studies, financial deepening of this type has generally been found to have a positive effect on medium-term growth [Beck and Levine (2004)]. Taken literally, this would suggest that the rise in banking scale over recent decades has provided a significant tailwind to medium-term growth in advanced countries. So it seemed in the pre-crisis period.

But that conventional wisdom has recently been called into question. IMF research has suggested that financial deepening can indeed be growth-positive – but within limits. There is a threshold at which private credit-to-GDP may begin to have a negative impact on GDP growth (Arcand et al. (2012)). That threshold is found to lie at a private credit-to-GDP ratio of around 80%–100%. This finding is consistent with earlier cross-country evidence suggesting that, at credit-to-GDP ratios above unity, output volatility tends to increase [Easterly et al. (2000)]4. This threshold lies significantly below current levels of financial depth in most advanced economies. In other words, taken at face value this evidence suggests that, at its current scale, banking could be acting as a headwind to medium-term growth.

Accompanying this dramatic rise in banking scale has been an equally dramatic rise in banking concentration. Figure 2 plots the evolution of the asset share of the three largest banks in the U.S., U.K., Germany and Switzerland over recent decades. The U.S. has undergone the most dramatic upward shift, with the share of the top three banks rising from around 10% to 40% between 1990 and 2007. For other countries, there is a less dramatic rise in concentration, but from a much higher starting point, with the top three banks accounting for between two thirds and three-quarters of assets in the U.K., Switzerland and Germany.

This dramatic rise in banking scale and concentration has been driven by financial liberalization. But it has also spawned an acute problem for society due to escalating expectations of state support for the banking system. These expectations generate lower funding costs, in particular for the largest banks, which in turn encourages further expansion and concentration, worsening the too-big-to-fail dilemma. There is a self perpetuating “doom loop” [Alessandri and Haldane (2010)].

The size of the resulting “implicit subsidy” from the state to the big banks has been the subject of recent study – and no little controversy [Noss and Sowerbutts (2012)]. There are a number of possible methods for estimating the subsidy. Perhaps the simplest is found by comparing the “standalone” and “support” ratings assigned to debt issued by large banks. The difference between these ratings gives an estimate, used by the market when pricing debt, of the probability of state support. Figure 3 plots the difference between these ratings for the 29 institutions deemed by the Financial Stability Board (FSB) last year to be the world’s most systemically-important. In the pre-crisis period, this difference averaged 1.3 notches. This sounds material, if not breathtaking. It suggests that, while non-trivial, too-big-to-fail may not have been a first-order driver of the rising scale and concentration in banking.

Yet even small notches of support can translate into big implicit subsidies if balance sheets are large. So it is for the world’s largest banks. Over the period 2002 to 2007, the implied annual subsidy to the world’s largest banks averaged U.S.$70b per year using a ratings-based measure (Figure 4). That is roughly 50% of the average post-tax profits of these banks over the period.

As the crisis struck, this implicit promise became explicit. Financial support was extended to the banking system in the
form of capital injections, guarantees and liquidity insurance. On some estimates, this support rose to three-quarters of annual GDP in some countries. In response to these interventions, there has predictably been a further ratcheting-up in ratings-implied degrees of state support to banks.

By 2009, the ratings difference had more than doubled to above three notches, with the implied monetary subsidy over U.S.$700b per year. That was well in excess of average annual pre-crisis profits of these firms. Even if an over-estimate, the scale of the implied subsidy signaled something dramatic was at play. Too-big-to-fail had become hard-wired into the structure and pricing of the financial system.

3. Systemic surcharges

One way of interpreting these implicit subsidies is as the market’s best guess of how much a policymaker would be willing to pay each year to avoid the failure of the world’s biggest banks. They proxy the expected social costs of big bank failure. In the jargon, they capture a systemic externality.

This notion of a systemic externality has underpinned recent academic and policy efforts to solve the too-big-to-fail problem. Brunnermeier et al. (2009) use this framework to motivate levying a tax – a “Pigouvian tax” – on institutions posing systemic risk externalities. This tax would be set at levels which offset the effects of the bank’s actions on wider society. A number of academics have since proposed measures along broadly Pigouvian lines (Archarya et al (2010)).

Rather remarkably, policy reforms in practice have followed closely in the spirit of these proposals. In 2010, the FSB announced its intention to introduce a “systemic surcharge” of additional capital on the world’s largest banks. In July 2011, the Basel Committee published a methodology for measuring global systemic importance based on indicators of bank size, connectivity and complexity, with additional capital of up to 2.5% depending on the scores for these indicators. This methodology was fully finalized in 2013.

In November 2011, the FSB endorsed this methodology and announced 29 global systemically important banks (G-SIBs). It has updated the list of G-SIBs on an annual basis since then.

---

5 The BofA Merrill Lynch indices are provided for informational purposes only and may not be used without the written permission of BofA Merrill Lynch. The BofA indices are not intended to be used for investment or other commercial purposes. BofA Merrill Lynch is licensing the BofA Merrill Lynch indices “as is,” makes no warranties regarding same, does not guarantee the suitability, quality, accuracy, timeliness and/or completeness of the BofA Merrill Lynch indices or any other data included in, related to, or derived therefrom, assumes no liability in connection with their use, and does not sponsor, endorse, or recommend (BOE), or any of its products or services.
surcharges will be phased in from 2016. Legislation is already in place in the E.U. to implement the systemic surcharge.

These proposals are clearly a practical step in the right direction. By boosting levels of capital in the system, the probability of big bank failure will be reduced. You would have got good odds back in 2007 that something as seemingly elliptical as a Pigouvian tax on systemic risk would have found its way onto the regulatory statute books. Now we have it. “We are all Pigouvians now,” even if most of us cannot spell it.

The practical question is how far systemic surcharges take us in tackling the systemic risk externality. In other words, at current levels, by how much have systemic surcharges reduced expected losses for the financial system? That is an empirical question. To assess it, consider the impact of capital surcharges on the expected losses facing the 29 institutions identified by the FSB.

Using each of these banks’ balance sheets, we generate a measure of default probability using the Merton (1974) contingent claims model. We assume that the base level of equity for these banks is 7% (the new Basel III minimum) and that, in the event of default, they suffer losses on their assets of around 30%. To keep things simple, banks’ assets are assumed to be (log-)normally distributed, in line with Merton (1974), and default occurs only when a bank’s capital is fully exhausted. Assume initially that default risks across banks are independent. This is a highly conservative assumption, as in practice bank default probabilities are highly correlated in times of stress. Indeed, bank correlation coefficients in crisis often head toward one. For that reason, this thought-experiment provides a lower bound on expected losses across the financial system.

Figure 5 shows expected losses across the 29 banks at different levels of the systemic surcharge. In the absence of any surcharge, expected losses across the system are just less than U.S.$200b per year. Were every large bank instead to face the maximum capital surcharge of 2.5%, then expected system-wide losses would fall by around 60% relative to their base level. And to remove 90% of the systemic externality – expected losses of around U.S.$5b – a surcharge of over 7% would be needed.

A more plausible experiment would be to assume a non-zero correlation among bank defaults. For example, the failure of a large bank, which caused it to fire-sale assets could impose externalities on other large banks holding these same assets (Wagner (2009)). To place an upper bound on expected losses in the face of these fire-sale externalities, assume instead a correlation coefficient of one. Figure 5 illustrates the impact on expected system-wide losses of a high degree of default correlation among the big banks. The expected system-wide loss increases to around U.S.$750 billion per year – similar in size to the implicit subsidy at its peak. A 2.5% surcharge now only reduces expected system-wide losses to around $350 billion per year. To lower expected system-wide losses to be around U.S.$5 billion per year would require a surcharge of around 15% – six times its current upper limit.

If anything, these thought-experiments probably produce conservative estimates of system-wide losses and the necessary systemic surcharge. For example, bank asset returns are in practice much fatter-tailed than the log-normal distribution. And in practice, banks are likely to default well before their capital is fully exhausted. Relaxing either assumption would push up

---

6 James (1991). We also assume bank asset volatility of 4% per year, in line with previous empirical studies.
estimates of expected losses and the surcharge necessary to curtail these losses [Schanz et al. (2012)]. Nonetheless, if expected system-wide losses are a reasonable proxy for the system-wide externalities large banks pose, this analysis delivers a rather gloomy prognosis. At current levels of the surcharge, a large chunk of the systemic externality would remain untouched. If too-big-to-fail is the problem, then systemic surcharges seem to offer only a partial solution.

4. Resolution regimes
Capital surcharges lower systemic externalities by lowering default probabilities for the world's largest banks. An alternative way of lowering those externalities would be to reduce the collateral damage associated with their failure. This has been a key motivation for a second strand of the reform debate – the design of effective resolution regimes.

Few examples better illustrate the costs of getting this wrong than the spiraling queues outside branches of Northern Rock in September 2007. Despite being a medium-sized retail bank, Northern Rock's failure caused systemic disruption and put taxpayers' money at risk. In response, the U.K. put in place in 2009 a special resolution regime for banks, providing the Bank of England with tools for winding down a failed bank.

As the crisis illustrated, financial failure which causes systemic disruption is not confined to banks. And to avoid taxpayer bailout, losses may need to be imposed on a wide class of bank creditors, including holders of debt as well as equity – for example, by "bail-in." Over recent years, these resolution lessons have been enshrined in banking legislation. For example, in the U.S. Title II of the Dodd-Frank Act was passed in July 2010. It creates a new regime for the liquidation of financial companies, banks and non-banks, which pose a systemic financial stability risk. It enables losses to be imposed on creditors in resolution, while also prohibiting state bailouts.

Internationally, in November 2011, the G20 endorsed the FSB's Key Attributes of Effective Resolution Regimes, developed by an international working group chaired by my colleague Paul Tucker. Efforts are now underway to align national resolution regimes with these principles. As part of that, in Europe, a draft Directive on bank recovery and resolution was published in June 2012.

These initiatives are an important practical step in the right direction, lowering the societal costs of bank failure. As with systemic surcharges, it is striking how much progress has been made in so short a space of time on so complex an issue. The practical question is how far this takes us toward removing the too-big-to-fail externality.

During a bank resolution, one way of ensuring continuity of banking services is by transferring assets and/or liabilities of a failing firm to a third party. But the only entity with sufficient financial and managerial resource to absorb a large asset or liability portfolio, without suffering chronic indigestion, is another big bank. So it was during the crisis – for example, Bear Stearns was swallowed by JP Morgan Chase and Merrill Lynch by Bank of America.

This makes for a rather uncomfortable evolutionary trajectory, with rising levels of banking concentration and ever-larger too-big-to-fail banks. Levels of banking concentration have risen in many countries since 2007, precisely because of such shot-gun marriages by over-sized partners. In other words, resolving big banks may have helped yesterday's too-big-to-fail problem, but at the expense of worsening tomorrow's.

One way of avoiding this problem is to re-capitalize a bank by bailing-in its creditors, rather than transferring its assets. But resolution rules of this type are not problem-free either. Like all policy rules, they face what economists call a time-consistency problem. Whether a rule is followed in practice depends on the balance of costs and benefits at the time crisis strikes, not at the time the rule is written. That is why policy might in practice lack consistency over time – hence time-inconsistency.

Consider that trade-off when a big, complex bank hits the rocks. On the one side is a simple, but certain, option – state bail-out. On the other is a complex, and less certain, option – resolution. Policymakers face a trade-off between placing losses on a narrow set of taxpayers today (bail-in) or spreading that risk across a wider set of taxpayers today and in the future (bailout). If governments are risk-averse and wish to smooth the pain across taxpayers and across time, then bailout may look attractive on the day.

Financial history certainly suggests so. The history of big bank failure is a history of the state blinking before private creditors
Recent crisis experience has written another chapter in this history. Next time may be different. For example, the public backlash against future bailouts could reinforce government’s resolve to impose losses on creditors. And recent U.S. legislation in principle locks the taxpayer cashbox and throws away the key. Looking forward, the issue is whether this ex-ante rule is ex post credible.

As Figure 6 illustrates, implied levels of support for the U.S.’s biggest banks are much higher than before the crisis. More telling still, the passage of Dodd-Frank appears to have had little impact on levels of implied state support. It is early days for this new resolution regime and credibility may take time to be earned. Nonetheless, at present, the market believes the time-consistency problem for big banks is as acute as ever.

Even if it might appear the expedient option on the day of crisis, it is questionable whether bailout is the optimal response over the medium term. Figure 7 looks at the response of bank and sovereign CDS spreads around the time of bank bailouts in a selection of crisis countries. While bailouts lowered bank CDS spreads, as might be expected, bailout came at the expense of a rise in sovereign CDS spreads.

It is not difficult to see why. The financial crisis has caused huge damage to the balance sheets of governments in advanced economies. For the G20 countries, the IMF forecast that the debt-to-GDP ratio will rise by almost 40 percentage points between 2007 and 2016, to almost 120%. At these levels, public sector debt may be a significant drag on medium-term growth [Rogoff and Reinhart (2010)].

For economies with large banking systems and without a credible resolution regime, this leaves policymakers caught between a rock and hard place. When the call comes to ride to the banking rescue, government may be unable to afford not to. But nor, at least over the medium term, can they afford to. This is just the dilemma facing advanced countries today.

---

[Disclaimer of liability notice. This may contain information obtained from third parties, including ratings from credit ratings agencies such as Standard & Poor’s. Reproduction and distribution of third party content in any form is prohibited except with the prior written permission of the related third party. Third-party content providers do not guarantee the accuracy, completeness, timeliness or availability of any information, including ratings, and are not responsible for any errors or omissions (negligent or otherwise), regardless of the cause, or for the results obtained from the use of such content. Third party content providers give no express or implied warranties, including, but not limited to, any warranties of merchantability or fitness for a particular purpose or use. Third party content providers shall not be liable for any direct, indirect, incidental, exemplary, compensatory, punitive, special or consequential damages, costs, expenses, legal fees, or losses (including lost income or profits and opportunity costs or losses caused by negligence) in connection with any use of their content, including ratings. Credit ratings are statements of opinions and are not statements of fact or recommendations to purchase, hold or sell securities. They do not address the suitability of securities or the suitability of securities for investment purposes, and should not be relied on as investment advice.]
5. Structural reform

One way of lessening that dilemma, and at the same making resolution and bail-in more credible, is to act on the scale and structure of banking directly. Perhaps not surprisingly, recent regulatory reforms have sought to do just that. They have taken seriously the maxim that, if a bank is too big to fail, then it is too big. The result has been detailed proposals for structural reform in a number of advanced countries.

In the U.S., the “Volcker rule” has been introduced. This prohibits U.S. operating banks from undertaking proprietary trading and restricts private equity activity. The rule was tabled in October 2011 and became law in July 2012. Banks have two years to comply. In the U.K., the proposals of the “Vickers Commission” include placing a ring-fence around retail banking activities, supported by higher levels of capital. The final version of these proposals was tabled in September 2011, and legislation to enact them is planned by 2015. Banks will have until 2019 to comply. Most recently in Europe, the “Liikanen plan” was announced in October 2012. It proposes that the investment banking activities of universal banks be placed in a separate entity from the remainder of the banking group. There are at present no plans to legislate these proposals.

As with the other reform strands, it is remarkable how quickly radical structural reform proposals are finding their way onto the statute book. And, although different in detail, these proposals share a common motivation: separation of certain investment and commercial banking activities. In theory, such a separation delivers financial stability benefits of two distinct types [Boot and Ratnovski (2012)].

First, separation reduces the risk of cross-contamination. Riskier investment banking activities, when they go wrong, can pollute and dilute the financial resources of the retail bank. This potentially inflicts losses (or fear of losses) on depositors. It may also constrain banks’ ability to make loans to the real economy when it might most need them. This is a crisis-time benefit of separation.

Second, separation can secure an improved pre-crisis allocation of financial resources from a societal perspective. High private return investment banking activities may crowd-out the human and financial resources devoted to high social return commercial banking activities. Investment banking activities might also piggy-back on the cheaper cost of deposit funding. In effect, universal banking allows privately optimal, but socially sub-optimal, cross-subsidization. This is a normal-times benefit of separation.

Both of these costs were evident ahead of, and during, the recent crisis. Ahead of crisis, resources gravitated to the investment banking side of the fence. Between 2000 and 2007, U.K. banks’ trading books rose six times as fast as their banking books. Human capital made the same journey, helped by investment banking salaries rising four times as fast as commercial bank salaries since 1980.

In the teeth of crisis, risk cross-contamination became a potent factor. Basic banking services in universal banks were often subject to severe disruption from trading book losses, which exceeded by many multiples the capital allocated to them. That is why national deposit insurance schemes were extended and, in some cases, became temporarily unlimited. It is also why repeated attempts have had to be made to resuscitate weak credit growth over the past few years.

So, how far will existing structural proposals take us in harnessing these benefits? Volcker, Vickers and Liikanen seek legal, financial and operational separation of activities. So, in principle, each ought to prevent cross-contamination at crisis time. Whether they do so in practice depends on loopholes in, or omissions from, the ring-fence. And each of the existing proposals has open questions on this front.

For example, the Volcker rule separates only a fairly limited range of potentially risky investment bank activities, in the form of proprietary trading. The Vickers proposals mandate only a limited range of basic banking activities to lie within the ring-fence, namely deposit-taking and overdrafts. And the Liikanen plans allow a wide range of derivative activity to lie outside of the investment banking ring-fence.

It could be argued that these loopholes are modest. But as the history of the Glass-Steagall Act demonstrates, today’s loophole can become tomorrow’s bolthole, today’s ring-fence tomorrow’s string vest. At a minimum, this suggests the need for full and faithful implementation of the spirit as well as the letter of the...
Volcker, Vickers and Liikanen plans, if risk cross-contamination is to be avoided.

A larger question-mark still hangs over whether these proposals will lead to a sea change in the allocation of resource to retail and investment banking. The cultures of investment and retail banking are quite distinct. Retail banking relies on forming long-term relationships, while investment banking is inherently shorter-term and transactional. Housing these sub-cultures under one roof makes achieving the necessary separation of cultures and capital a significant operational headache.

At a minimum, such a separation of culture and capital is likely to require entirely separate governance, risk and balance sheet management on either side of the ring fence. Without that, human and financial resource allocation to either side of the ring-fence will become blurred. For example, without separate debt issuance for retail and investment banking, the cost of debt for a big bank will be a blended mix. The implicit subsidy in funding costs would then remain, and with it one of the main distortions associated with too-big-to-fail.

Only time will tell whether cultural separation can be achieved under the existing structural reform proposals. In the go-go years, will these reforms be sufficient to prevent the grass always appearing greener on the riskier side of the (ring-)fence? This is the acid test of the structural reform agenda.

6. Where next?

Progress has been made over the past few years toward eliminating too-big-to-fail, with further progress on implementation planned. But today’s task is even more daunting than before the crisis. The big banks are even bigger. The system itself is more concentrated. And despite reform efforts, the market’s best guess today about tomorrow’s implicit subsidy is far larger than before the crisis, at over U.S.$300b per year (Figure 3). The market believes that illicit state promise is even more likely to be kept.

The wrong conclusion to draw would be that existing reforms have failed or are unnecessary – quite the contrary. Rather it is that these reform initiatives, while necessary, may be insufficient to eliminate the too-big-to-fail externality. If so, what are the alternatives? Several have been mooted.

Resizing the capital surcharge is one possibility. This would further reduce default probabilities among the biggest banks, thereby lowering the expected system-wide losses associated with big bank failure. Taking the earlier illustrative example, to reduce materially expected system-wide losses for the world’s largest banks would require a capital surcharge several times larger than its current upper limit. Interestingly, this would take bank capital ratios to levels not dissimilar to recent quantitative estimates of their optimal value (Miles et al. (2011), Hellwig et al. (2011)).

Placing limits on bank size is a second option. By reducing balance sheet exposures, this measure would reduce directly system-wide losses in the event of big bank failure. The Dodd-Frank Act includes an explicit limit on the maximum deposit market share of U.S. banks, capping it at 10%. But this does not prevent banks rising to a scale, relative to GDP, at which they could imperil state solvency. For that reason, limiting bank size relative to GDP has recently been proposed by a number of commentators and policymakers (Fisher (2011), Hoenig (2012), Johnson and Kwak (2010), Tarullo (2012)).

Full structural separation of investment and commercial banking. A modern-day Glass-Steagall Act, has continued to attract support. The main benefit this would bring, relative to structural ring-fencing, is that it would eliminate loopholes from the ring-fence and better ensure that the distinct cultures of retail and investment banking were not cross-contaminated. That would lessen the risk of basic banking activities being starved of human or financial capital, both ahead of and during crisis. Full separation may also be operationally simpler to implement than the existing structural proposals.

Finally, enhanced banking competition would potentially help to reduce some of the problems of too-big-to-fail by reducing the degree of banking concentration. Greater exit from banking through enhanced resolution regimes can help. But, a bigger problem still is bank entry: the U.K. went 100 years without a new retail bank being set up. One potential barrier to banking entry is the difficulty of switching deposit accounts and loan contracts. A shared banking platform, containing customer account details, would dramatically reduce the frictions in search and switch for deposit and loan products for customers (Leadsom (2012)). It could lower materially barriers for new banking entrants.
A powerful counter-argument to all of these more radical proposals is that they could erode the economies of scale and scope associated with large banks. These economies clearly do exist in banking, as they do in other industries. For example, fixed costs are large in finance and spreading them widely ought to deliver productivity improvements.

The interesting question is at what point these economies of scale are exhausted. Indeed, informational and managerial diseconomies of scale are likely at some scale, whatever the business line. In his classic theorem, Ronald Coase tells us that firms will seek a privately optimal size which balances the benefits of economies of scale against these diseconomies [Coase 1937].

How does all of this apply in banking? The empirical evidence on economies of scale and scope in banking is surprisingly patchy. Early studies, using data from the 1980s, failed to find scale economies much beyond bank asset sizes of around U.S.$100m [Pulley and Braunstein (1992)]. Empirical studies in the 1990s nudged up the optimal bank scale to around U.S.$1 billion [Amel et al. (2004), Mester (2005)].

Most recently, a small number of studies using data from the 2000s have pointed to scale-economies at much higher asset thresholds. For example, Wheelock and Wilson (2012) find scale economies for banks with assets up to U.S.$1 trillion and Feng and Serilitis (2009) for bank with assets up to around U.S.$1.5 trillion. Using data on banks with assets in excess of U.S.$100b, Mester and Hughes (2011) not only find scale economies, but argue that these may increase with bank size.

At face value, these findings pose a real challenge to policy options which re-size banks. Or do they? Bank of England research has re-looked at the evidence on economies at different banking scales [Davies and Tracey (2013)]. They show that in standard models, there is evidence of scale economies for banks with assets above U.S.$100b. Indeed, these economies tend to rise with banking scale.

But this finding is based on estimates of banks’ funding costs which take no account of the implicit subsidy associated with too-big-to-fail. Removing this subsidy raises banks’ funding costs, lowers estimates of bank value-added, and thereby reduces measured economies of scale. Davies and Tracey (2013) show that once an allowance is made for the implicit subsidy, the picture changes dramatically. There is no longer evidence of economies of scale at bank sizes above U.S.$100b.

This evidence reconciles Coase’s Theorem with the too-big-to-fail phenomenon. In line with Coase, banks have chosen the size that maximizes their private value. But implicit subsidies may have artificially boosted the privately optimal bank size. Subtracting this subsidy, removing the state crutch, would suggest a dramatically lower socially optimal banking scale. Like King Kong and Godzilla, these giants would arguably then be physiological impossibilities.

What about economies of scope? A recent study by Boyd and Heitz (2011) conducts a simple, but compelling, thought-experiment. They compare the lowest-available estimate of the social cost of the crisis with the highest-available estimate of the private benefit of scale and scope economies in banking. The social costs of too-big-to-fail exceed the private benefits of scale economies by an order of magnitude.

7. Conclusion
In his 1928 article, J B S Haldane observed that when “you drop a mouse down a thousand-yard mine shaft it walks away, a rat is killed, a man is broken, a horse splashes.” When big banks disappeared down the mineshaft in 2008, their splashes generated a tsunami. To prevent that, their physiology needs to change. Existing change initiatives are right in direction, but may be insufficient in degree. There may be a distance to travel before banking is the right size.

Commentary by Philip Booth: Is big dangerous?
Andrew Haldane begins his paper with a lesson from biology. I think that there are other lessons we can learn from evolutionary biology as well as the one cited by Andrew. The family of the economist F. A. Hayek, whom I shall quote later, were also biologists and Hayek was very familiar with the subject himself. Hayek drew analogies between the market process and evolution. A market process is a process of evolution; of trial and error; of the copying of good ideas; and of the failure of bad ideas. This requires that banks fail. Clearly to ensure that we have an orderly market, they must fail in an orderly fashion. But, failure is essential if competition is to be functional.
The second lesson that Hayek took from the biological sciences is the way in which markets were extended networks of individuals and institutions in which entrepreneurs would respond to the information that was relevant to them. That information cannot be centralized. As such, the detailed regulation of market and the replication of market processes by government agencies, is not just difficult or undesirable, but impossible.

I will come back to these issues in a moment, but first I want to focus on the specific issues brought up in Andrew's paper. The paper focuses, to a large degree, on the problem of too-big-to-fail banks. It is argued that these have a large implicit subsidy from the taxpayer — and I entirely agree with that. It was mentioned that certain reforms have been put in place and possible further reforms are then suggested by Andrew.

However, it is worth noting that it is not entirely clear that a concentrated banking system — at least before the regulatory reforms that we have seen in Britain — was more dangerous than a less concentrated one. If you compare the U.K. and the U.S. banking systems with each other over the last century or so, the U.K.'s has consistently been extremely concentrated and that in the U.S. has not. Arguably, the U.K.'s has been more stable — certainly that was the case during the Great Depression. Also, the failed savings and loans sector in the U.S. was not a concentrated sector either, but that demanded a huge bailout. I think there are three nuances that we have to add to Andrew's argument:

1. In the traditional analogy between the payments' system and the plumbing system, the system breaks at its weakest link — not necessarily its biggest link.

2. If similar mistakes are made throughout the system by market practitioners for whatever reason then the impact (as in the savings and loans crisis) could be just as great as if one big bank failed in a concentrated system.

3. In a concentrated banking system it can be easier to coordinate a response to the failure of one link in the system. Indeed, the remaining banks may have a strong incentive to coordinate a “lifeboat” without state help because the externalities of failure are more easily contained within a small group of institutions that can coordinate effectively.

In addition to this, larger banks might be more diversified, and therefore less prone to failure. We have, it seems, implicitly accepted that “bigness” and/or concentration go together with danger in the banking system because of the events of 2008, but this has not always been the experience.

Andrew's point about concentration and size, though, is more pertinent under the new regulatory regime. Under that regime, a resolution process to ensure orderly failure is the key. Our legal system for dealing with a failing bank in 2008 was not fit for purpose. Therefore, the failure of any bank could cause chaos (even the failure of a small bank). If that legal system becomes fit for purpose, it will become much easier to resolve smaller and simpler financial institutions and big, while complex institutions could still cause problems.

**Is structural change and capital regulation the answer?**

But, what else should we do if we accept the argument that there may be a residual problem of too-big-to-fail? I struggle with the idea of imposing structural change through legislation. It is a very blunt tool. I think that The Vickers' approach is better than the E.U. proposals and Dodd Frank, but I still have some concerns about Vickers' proposals and, indeed, the government has relaxed the rules proposed by Vickers in specific cases.

Further, capital regulation is another possibility and we will, of course, have even more capital regulation under the new regime. However, I reject that approach entirely for all the reasons that Andrew laid out in his “Dog and frisbee” speech at Jackson Hole [Haldane and Madouros (2012)]. We must get away from discretionary regulation designed to ensure that each bank has exactly a 0.5% probability of failure — or whatever the probability is. Simple capital rules get gamed and lead to complexity. Complex rules get gamed using complex methods and lead us to where we are now. Last year, there were 14,200 new banking regulations worldwide; before 2008, there were probably 1,000,000 paragraphs of FSA banking regulation; and the Dodd Frank bill will probably be 30,000 pages long with the associated regulations. This is completely crazy and also leads to the system being captured by a combination of the industry and a bureaucratic clique. Of course, any system of capital requirements designed to prevent failure also prevents competition from operating properly. Competition needs failure.
International capital regulation is even worse as it institutionalizes systemic risk: if the regulations incentivize particular types of behavior, the whole system is liable to go wrong at the same time. And this leads to an important question: if we are convinced that the resolution regimes will work, then why do we need a capital regime to stop failure happening? If we are not convinced that our resolution regimes will work, we should think again.

At the same time, for all the reasons we understand from public choice economics, and for the reasons laid out by Hayek, regulatory discretion — which is an alternative to detailed rules — is very dangerous. As Hayek (1989) said: “The curious task of economics is to demonstrate to men how little they really know about what they imagine they can design.” Even the most well-meaning and intelligent regulators should not be trusted with large amounts of regulatory discretion.

Stable legal frameworks: the E.U. elephant in the room

It is because Andrew understands this that his paper has discussed other possible solutions to the problem of too-big-to-fail. They may have some merit; but I worry that they would not last more than a generation as developments came along that transcended their relevance. So, I would like to add one more possibility to the mix. Indeed, I would like to make a proposal that could make all other banking regulation—except disclosure requirements—irrelevant.

I have come to the view from comparing systems of financial regulation in different sectors at different times in history that there is only one stable, just and effective solution to these problems. Everything must depend on the ability to ensure that a financial institution fails safely; and regulatory discretion must be extremely limited. Andrew partly accepts this starting point. However, in trying to achieve this, I think the focus on the separation of investment and trading and retail banking has been misplaced. There are too many grey areas, and an inherently incorrect assumption has been made that retail banking is somehow not risky. Separation of retail and trading functions may sometimes be a good idea, but it will not be necessary in some cases and will certainly not be a panacea.

In looking at the retail/invest/trading divide we have missed another aspect of bank structure that makes resolution much more difficult: international branching, especially when legal frameworks, deposit insurance schemes and the legal personality of subsidiaries do not match up. Perhaps nobody dares to speak out about this because of issues regarding E.U. regulation and its jurisdiction over U.K. law that would be raised. But, many financial firms do not use branching for their international operations and work through subsidiaries. There is an active international insurance market operating through subsidiaries too — even in the E.U. where branching is encouraged. You can have free trade without banning — as the E.U. does — requirements to have subsidiaries in different jurisdictions. It is often the international branching of banks that makes resolution so difficult — the Icelandic case being one important example. I would not go as far as requiring international subsidiarization of banks, but it ought to be an option that is available.

Specifically, I think that the most stable approach would be as follows. The legal duty of the Bank of England in the field of banking regulation should be to ensure that any bank can be wound up safely without it disrupting the rest of the banking system, and without undue calls on the deposit insurance system. The Bank of England should not enforce capital requirements to prevent failure, but should have the duty to ensure that failure is orderly if it happens.

This, of course, requires power. If it believes that a bank could not be wound up in an orderly fashion, the Bank of England should have the power to enforce structural change to ensure that it can be wound up so that the corporate structure is simplified, subsidiaries are set up and/or living wills are drawn up. This could involve separation of international business into subsidiaries, separation of trading and retail activities and so on, but it may not necessarily do so. I would not regard this as regulation as such. The Bank of England would simply be placing requirements on those banks to which it provided lender of last resort support that were deemed necessary in order for banks to have the principles of their special bankruptcy law applied to them. This would be complemented by depositor preference and risk-based deposit insurance levies, if deposit insurance remains.

Obtaining a banking license, therefore, would no longer depend on whether a bank was safe but on whether it could fail safely. This certainly would facilitate new entry as new entrants would hardly need to be regulated at all. There would, of course, need to be a right of appeal to an independent technical panel and then to a court, in case the Bank of England was excessive in exercising these legal powers, but the judge would decide on the basis of a
simple principle – what structure is necessary to ensure that a bank can be wound up so that as many as possible of its contractual obligations can be met in an orderly way.

Those systems of financial regulation that have succeeded in the past have relied on good legal principles and the incentives that the market itself provides. In my view, there should be less emphasis on regulatory control, preventing failure and on structural reform. I have considered lots of alternatives and I cannot think of another enduring and stable way of resolving the problems that Andrew has talked about in this paper as well as in his dog and the frisbee speech. We must retreat from detailed regulation of all behaviors to simple legal frameworks. I should add that this leaves unconsidered whether regulation is necessary for consumer protection purposes. In my view, it is not, but that is a separate argument.

In conclusion, if we really do have a credible failure regime then we do not have to worry about whether the socially optimal size of banks is different from the privately optimal size or whether socially optimal capital levels are different from privately optimal levels. If we get the legal framework right, a lot of the problems with which we struggle are no longer relevant. I hope that Andrew might add this proposal to his list of possible ways of dealing with the too-big-to-fail problem.

References

Avery, S., and J. Haltiwanger, 2011, “A new look at the job creation and destruction literature,” mimeo
Davies, R., and B. Tracey, 2013, “Too big to be efficient? The impact of too-big-to-fail factors on scale economies for banks,” mimeo
Fisher, R., 2011, “Taming the too-big-to-fails: will Dodd-Frank be the ticket or is bank-lab surgery required?” Remarks before Columbia University’s Politics and Business Club
Wagner, W., 2009, “In the quest of systemic externalities: a review of the literature,” mimeo
Wheelock, D. C., and P. W. Wilson, 2012, “Do large banks have lower costs? New estimates of returns to scale for U.S. banks,” Journal of Money, Credit, and Banking, 44(1), 171-199

The Journal of Financial Perspectives 25
Part 1: Strategic

Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives

Paula Jarzabkowski
Professor of Strategy, Cass Business School, City University

Rebecca Bednarek
Research Fellow, Cass Business School, City University

Laure Cabantous
Senior Lecturer, Cass Business School, City University

Abstract
In recent years, the structure of the reinsurance industry has evolved rapidly in response to, among other factors, changes in the nature of risks to be insured (e.g., climate related risks), technological innovations (e.g., Cat models), the globalization of the financial services industry, and the concentration of the primary insurance industry. Drawing on a global qualitative dataset of interviews and observations in all the main reinsurance markets, this article provides an overview of the changing dynamics in the reinsurance industry and highlights some of the implications of these changes for reinsurers. We first provide a general overview of the changing structure of the reinsurance industry, highlighting its increasingly competitive nature and why these changes might exert pressure on reinsurer’s profits. Second, we focus on consolidation in the primary industry as a central industry dynamic. Third, we show how this impacts the types of reinsurance products cedents want from the reinsurance industry. Finally, we outline some implications of this for reinsurers, identifying matters of both strategic and operational concern for them.
Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives

Introduction
The modern reinsurance industry, which is basically the insurance of insurers, has a long history, at least in Western Europe, with the first reinsurance treaties arising in Germany after 1820 and the first reinsurance company, Kolnische Ruchversicherungsgeselschaft (Cologne Re), founded in 1842 [Kopt (1929)]. Historically, there has been little change and a relatively slow rate of innovation in the reinsurance industry [Tuohy (2008)]. More recently, however, this mature industry, which is part of the global financial service industry, is experiencing intensified competition and regulation. In particular, as a result of consolidation in the primary insurance industry, cedents (i.e., the insurers who look to buy risk cover from reinsurers) are changing the way they buy reinsurance, and these changes are escalating rapidly with many implications for reinsurers. In this article, we explore the increasingly competitive and changing dynamics of the reinsurance industry, and the implications of these dynamics for reinsurance executives. To do so, we draw from a novel qualitative study of the global reinsurance industry, which covered the industry’s main players and markets (see textbox 1).

Our analysis of the structure of the reinsurance industry, which is based on Porter’s five forces framework, first shows why it is more difficult to make money in the reinsurance industry than ever before [Porter (1980)]. Second, we conduct an in-depth analysis of the industry life cycle of the primary insurance industry. Third, this lifecycle analysis reveals how consolidation in cedents’ buying of reinsurance is evolving the very notion of the reinsurance product as cedents’ needs shift to a new class of product. Such changes are evidently hard for most reinsurers to respond to and keep abreast of. We will, therefore, end with the implications of such trends for reinsurance executives. We highlight implications for different types of reinsurers as well as specific issues in relation to the internal operations of reinsurers, the specific issue of non-traditional reinsurance products (Alternative Risk Transfer) and reinsurance needs in emerging markets.

General overview of the reinsurance industry structure
A review of the reinsurance industry suggests that the “risk business” is getting riskier and more competitive. To structure the findings of our empirical research we use Michael Porter’s (1980) “five forces” framework. Porter’s model is a tool derived from industrial economics that is widely used for analyzing the attractiveness of an industry (i.e., its potential to generate higher than average profit margins) and for classifying features that affect the degree of competition in the industry. It identifies that industry structure, and its associated profit margins are determined by the interplay of five forces: bargaining power of buyers, bargaining power of suppliers, threat of new entrants, threat of substitute products, and intensity of industry rivalry. In the reinsurance industry, our analysis suggests that the profitability in the non-life reinsurance segment (our focus here) is eroding [Datamonitor (2011)].

First, and foremost, is buyer power. In reinsurance terms, insurers pay to “cede” or transfer risk to reinsurers, and these insurance-industry buyers are therefore known as “cedents.” As will be discussed in more depth below, the general trend for insurers is consolidation, meaning bigger and more powerful cedents. Of course, buyer power is not uniform; smaller insurers and insurers in emerging markets remain highly reliant on reinsurance as a proxy for capital, freeing up internal capital reserves to fund growth. But considered as a group, cedents have been bulked-up by consolidation into large, often global, insurance companies that are “heavy hitters” in their negotiations with reinsurers. Large cedents have massive market capitalizations that can be far superior to those of large

The primary data for this study was conducted over three years from 2009 to 2012. It consisted of:

1. 22 reinsurance firms, 3 brokerage firms and 36 insurance firms across 17 countries, and 61 offices in all non-life lines of business
2. 837 observations of multiple interactions and periods of analysis
3. 446 interviews (reinsurers, brokers and cedents)

Following 2012, extensive secondary data, for example, press articles and reports, and conversations with industry, for example, meetings with senior executives and strategy Masterclasses for reinsurance professionals continue.

Textbox 1: Global reinsurance study
Source: Bank of England calculations
reinsurers. For example, Allianz, a major European insurance firm, had a market capitalization of €53.4 billion (in August of 2013), its reinsurer Munich Re’s market capitalization was €24.4 billion and Hannover Re’s was €7.1 billion\(^1\). These large cedents also place billions of dollars of insurance, and thus dwarf even the largest reinsurers in terms of premium volume; while AXA as one of the largest insurers in the world underwrote U.S.$108 billion in premium in 2010; Munich Re, as the largest reinsurer, underwrote “only” U.S.$29.3 billion (Reactions (2012)). Such large cedents can, therefore, use their buyer power to lower the rates they pay for the amount of reinsurance cover they buy. Ultimately, they can afford to reduce their cover as their size allows them to retain more of their risks in-house (see below). Thus, in general, consolidation of cedents is squeezing reinsurance profits.

**Substitute** products are threatening if they offer a better price-performance ratio than an industry’s core products. In many segments of the global reinsurance market, for over a decade cedents have been using methods of Alternative Risk Transfer (ART), primarily in the form of Cat bonds, as a cheaper substitute for traditional reinsurance products [Loubergé, et al. (1999), Bougen, (2003)]. Cat bonds are insurance-linked securities for which returns depend on the occurrence of a specific insurance event. They are not normally considered part of the core reinsurance market, as they are not generally offered by reinsurers, but rather by hedge funds and pension funds. In fact, since 1996, the Cat bond market has seen U.S.$44b of cumulative reinsurance issuance. While Cat bonds, together with the hedge funds and pension funds that provide them, took a hit in the financial crisis, they have since bounced back and, by 2012, reached their highest level for new issuances and outstanding volumes in four years [Aon Benfield (2012a)]. All of this is money that is not going to reinsurers via traditional reinsurance products. Furthermore, importantly Cat bonds usually compete favorably on price with traditional reinsurance products, placing further competitive pressure on reinsurers.\(^2\) Another important form of substitution is internal retention: cedents simply doing without the reinsurance product by instead covering in-house more of their own risks. This option is in fact connected to the increased buyer power described above. As one chairman of a large cedent outlined: “In the long term if you understand risk capital and how it works then you keep more of your risk yourself. But, no doubt, the long-term consequence is to reduce the amount of reinsurance that you place. In our case, we placed a fifth less than we did just five years ago.” (Cedent, interview) For instance, in-house retention of casualty and property-per-risk is increasing as cedents increasingly consider that they can carry more of these risks themselves, rather than transferring them to reinsurers [Aon Benfield (2012b)]. In general, use of substitute products, and internal retention, both of which deprive reinsurers of premium, are increasing as a result of consolidation in the insurance industry.

Reinsurers are reliant on various types of suppliers, such as suppliers of capital (e.g., capital inflows from investors within the broader financial market), suppliers of analytic services (e.g., risk modelling companies), suppliers of other services (e.g., brokerage firms) and the general supply of human resources – well-trained, highly skilled staff. If the cost of capital, services or human resource goes up, it eats into reinsurance profitability. To provide two examples, let us consider the power of brokers (the intermediaries or suppliers of clients to reinsurers) and modelling companies. First, in the past years, brokers have consolidated alongside their buyers (see above), so attaining global scale and scope. As a result, three main global brokers, Aon Benfield, Guy Carpenter and Willis, now have 85% market share of the industry.\(^3\) These changes have resulted in a largely oligopolistic broking market. As these brokers have grown in size, the reinsurance industry has become increasingly “brokered” and previous “direct” (non-brokered) markets such as Europe increasingly depend on brokers in a way that was unthinkable 15 years ago. New players use brokers to access new markets and previously-direct reinsurers are also increasingly relying on them. Further, as these large consolidated brokers have grown in strength, they are more able to provide services to their clients (such as analytics) which large reinsurers used to provide as part of their differentiation, so increasing competitive pressure on reinsurers.

---


\(^2\) For example: http://www.artemis.bm/blog/2013/07/12/le-and-collateralized-capital-flow-benefits-reinsurance-buyers-at-renewals-aon/

\(^3\) Source: http://www.holborn.com/%5CData%5CSites%5C1%5Cpdfcontent%5CBestsReview-HoldingTheirOwn.pdf
Second, modelling companies form another important group of suppliers to the reinsurance sector. Even if reinsurers have their own internal models, they are reliant on three big modelling companies, AIR, RMS and EQE\(^4\), due to the widespread usage in the industry of these few select models. Modelling companies are suppliers for the industry as a whole, providing a framework for analyzing risk which brokers and cedents – as well as reinsurers’ competitors – use. The power of these modelling companies has increased as the industry has become more technical in the last decade. This is partly in response to the demands of regulation; and partly because of cedents’ bundling of analytically challenging “super-risks” (discussed below). In general, increasing technical sophistication is giving more power to suppliers of analytical models, and to potential reinsurance employees with rare technical skills; both drive up the cost base for reinsurance operation.

Like substitute products, new entrants compete on the basis of price-performance ratios, but as industry outsiders they have to overcome barriers to entry. In reinsurance, close longstanding relationships between reinsurers and cedents have traditionally posed an important barrier to entry; but recently, their perceived value to cedents seems to have waned. The regulatory environment represents another such barrier to entry due to its (increasingly) stringent nature (Directive 2009/138/EC of the European Parliament and Council). While these barriers offer some protection for incumbent reinsurers, they are far from insurmountable especially from among hedge funds and insurers. Such companies can more easily scale entry barriers, using their prior knowledge and resources. As reinsurers explained: “Hedge funds – we see them as a threat because they take the steam out of our market. Smart money knows when there is money to be made. So when profitability is high, “Money” goes and starts a new Bermuda entity, and that takes a little bit of the steam out of our market” (Reinsurer, interview). In particular, hedge funds have been providing both CAT bonds (as described above) in the reinsurance market and also traditional reinsurance products. Consequently, the two trends of new entrants and substitute products are closely connected. With their established pools of investment capital, they jump the barriers to entry represented by the capital intensity of reinsurance. In particular, there was a strong increase in the number of such new entrants following the 2005 hurricane season (which included Katrina), when rates where at a historical high, a phenomenon referred to as the “new Bermudian market.”

In 2011–12, several well-known hedge funds continued to announce the formation of reinsurance companies in response to fickle investment opportunities elsewhere, and because they wanted to use reinsurance premium as a captive pool of capital to reinvest in the hedge fund itself. Examples include two U.S.$500m Bermudian reinsurance companies, TP Re and SAC Re, set up by hedge funds in 2012.\(^6\) This trend continues to increase: 2013 saw the largest ever Cat bond issued to a first

---


time sponsor, a U.S.$750m issue for Citizens Property Insurance (Florida). This was a company that many thought would never tap the Cat bond market; but for Citizens the benefits were clear: “they have upped their overall level (of cover) ... at a lower rate than what Citizens paid in the reinsurance market last year.”

Increased rivalry. As Figure 1 shows, our analysis suggests that every one of the five forces has increased its pressure on reinsurance profits over the past couple of years, and led to intense rivalry between reinsurers. As one reinsurer states of this recent increased competitiveness: “There has been a big change from an almost guaranteed profit ... to much more capital flowing in to the industry, and really leading to new products and much increased competition. At the end of the day, it has become much more difficult to make money with reinsurance” [Reinsurer, interview]. In particular, this is because many of these dynamics are interrelated – for example, cedent power, ability to retain risk rather than cede it as reinsurance premium, and an increased ability by those powerful cedents to engage with new capital providers, such as hedge funds, to provide new products, such as Cat bonds. In the following section, we explore one of the primary sources of this increasing pressure.

The insurance industry life cycle: a mature industry facing increasing concentration

Trends in the primary insurance industry are likely to change what insurance companies need from their reinsurers. We now discuss the insurance industry life cycle [Klepper and Miller (1995), Peltoniemi (2011)], provide evidence of consolidation in this industry and show how this consolidation impacts the products reinsurers are required to deliver.

While in some developing markets, such as those in parts of Asia, Latin America and Eastern Europe, the number of insurance companies is growing [SwissRe (2011)], the overall trend in the insurance industry is one of consolidation into a small number of key players [Connings (2013); Applebaum (2012)]. For example, in the U.S., the 10 largest Property and Casualty insurance companies now control 50% of the entire U.S. insurance market; and in auto insurance, that share is held by just five companies [Applebaum, (2012)]. This consolidation is consistent with the industry life cycle analysis, which predicts that industries heading toward maturity, such as the insurance industry – are characterized by the shakeout of small players, as Figure 2 shows.

Over the last 10 years, large insurance companies have sought to enhance their product and geographic reach, client base and operating efficiency (i.e., economies of scale and synergies) through a range of organic growth, joint venture and M&A activity. For instance, both the number and the value of M&A activities have increased in the insurance industry over the last decade [Deloitte (2012a)]; and in December 2012 alone, 75 deals worth a total of US$18.25b were reported. As powerful insurance companies have grown, and moved to the new and more profitable emerging markets, smaller insurers have protected their bottom lines by divesting non-core or underperforming businesses and subsidiaries, while withdrawing from foreign markets where they lacked sufficient scale [Deloitte (2012b)]. Such growth in leading insurers indeed correlates with a decrease in the overall number of players in mature markets.

---


Changing competitive dynamics in the reinsurance industry: implications of changes in buyer behavior for reinsurance executives

For example, the number of companies in Europe fell by almost 3% in 2011, with the number in some more crowded markets, such as the U.K. dropping by as much as 8% in a single year [Insurance Europe (2013)]. The “lost” companies failed in the face of competitive pressure, merged, or were acquired. The impact of such consolidation is evident, with large players now dominating insurance markets. Their presence is felt through their size, as evidenced by their massive market capitalization, and their scope in operating across multiple markets. In short, over the last 10 years (between 2002 and 2011), on average, leading insurance companies increased their total revenues by approximately 37%. Further, 20 of the top 25 companies increased their asset size in 2011 [AM Best (2012)].

The global consolidation of insurers, resulting from the fact that this industry is now mature, changes cedents’ strategies for buying reinsurance. In the following sections, we discuss how industry life cycle effects cascade from the primary insurance industry to affect the reinsurance industry.

The implications of cedent consolidation for reinsurance buying
Our research shows that insurers do not all want the same type of reinsurance product from reinsurers. Depending on their size for instance, insurance companies have different priorities in the purpose, products and organization of reinsurance buying. Specifically, we identified differences in the reinsurance needs of four central types of cedents (see textbox 2). This means that as the general industry trend moves away from local market players and becomes dominated by global and regional players the reinsurance products that reinsurers are required to deliver also change.

What determines cedent buying preferences?
From our research, we found that cedents’ reinsurance buying needs (or preferences) depend on two main criteria: their capitalization and the need for central coordination. These depend on size and scope of the insurer.

Textbox 2: Insurer-types

1. Emerging market insurers are those that operate in a single emerging market territory, or in a small number of similar territories and require reinsurance for access to capital and to alleviate overall portfolio volatility.
2. Local insurers are those that, while well-established within well-known territories, still practice largely within their country or state of origin.
3. Regional insurers are those that have extended beyond their domestic market to include surrounding regions.
4. Global insurers are at the peak of all three dimensions, as they operate in diverse territories and cover different lines of business through their complex multidivisional structure.

- Capitalization – the purpose of reinsurance is to support risk transfer through capital provision. In general, insurers need to hold capital reserves in order to cover potential payouts to their policyholders in the event of losses. Reinsurance transfers some of the primary-insurers’ risk to reinsurers who, in return for a premium payment, hold their own capital in reserve for large payouts to multiple policyholders in the case of an event that causes a particularly large loss. This enables insurers to cover more clients without holding all the capital themselves or free up some of their capital to invest in furthering their growth ambitions. Insurers’ level of capitalization influences their reinsurance needs. Less well-capitalized (i.e., smaller) players require reinsurance as a source of affordable capital, enabling them to grow or to cover greater risk for their clients. By contrast, well-capitalized insurers, particularly those with high capital efficiency arising from well-diversified portfolios, do not need reinsurance to either grow or to transfer some of the risk from their portfolio. Rather, they require it to cover peak risk – those “Armageddon scenarios” where a single event, such as a hurricane, or the asbestos crisis, wipes out a major part of the portfolio, and the company is flooded with claims. The growth of such large well-capitalized players arising from the consolidation of...
outlined above is, therefore, changing their reliance on reinsurance cover and the types of products they require.

- **Need for coordination** – the need for coordination is a function of the insurer’s size and scope. As insurance companies grow into new markets, they evolve from allowing local operating companies (LOC) to purchase their own local reinsurance cover to coordinating buying centrally across all LOCs. This central coordination of reinsurance purchasing enables capital efficiency through diversification, avoids duplication regarding the reinsurance being bought as a firm acquires more LOCs, and ensures that a group has oversight of, and is adequately hedged for, risks taken by LOCs. By contrast, smaller companies have fewer opportunities for capital efficiency and less need for formal coordination of reinsurance buying. Globally coordinated purchasing of reinsurance is also a more efficient and less costly working practice. Consequently, with the growth of global cedents, there are more opportunities for coordination – globally and regionally – whereby risks across multiple LOCs worldwide or regionally, such as aggregated catastrophe covers, can be “bundled” together. This will now be discussed.

**The rise of super-risks and alternative risk transfer products**

As discussed above, the types of reinsurance products that cedents need depend on their capitalization level and their need for coordination, which are, in turn, a feature of their size. The reinsurance industry has, therefore, become dominated by products that align with large, well-capitalized and centrally coordinated global buyers’ changing reinsurance needs. Specifically, our research identified three main reinsurance products that are associated with different types of cedents’ needs:

- **Bouquet products** bundle multiple or heterogeneous lines of business all related to a single territory and are usually associated with quite small, local and emerging market buyers and small insurers.14

- **Standalone products** cover one line of business, for example a single type of risk, such as a third party motor liability product, within a clearly defined territory or closely related territories, and are usually associated with local buyers that have grown to be significant in their local markets but have little business outside these markets.

- **Super-risk products** bundle reinsurance cover from a homogenous line of business, such as catastrophe risk, across multiple perils, such as wind and flood, and multiple territories; increasingly globally. These products are usually associated with global and regional buyers who want to centralize their buying.

Our analysis shows that, as a result of consolidation, some of these traditional reinsurance products are declining, while those that meet the needs of large insurance companies are increasingly dominant and command the majority of premiums in the reinsurance industry.

---


13 Line of business is the general classification of type of (re)insurance being written, i.e., homeowner, aviation and marine, among others (http://www.guycarp.com/content/guycarp/en/home/the-company/media-resources/glossary/l.html).

14 When insurance companies are very small and it is therefore not worthwhile for reinsurers to reinsure their portfolio as standalone programmes they typically bundle different lines (e.g., marine combined with property) into a single reinsurance product, the bouquet. Demand for a bouquet can also be connected to not having enough statistics on the various parts of your portfolio to place on their own.
Implication 1: Declining bouquet purchases and standalone products

Given the industry trends outlined above, demand for these “bouquet” and “standalone” products will decrease or remain stable in the coming years. Linked with very small, local and emerging market buyers, these bouquet products or “bundled heterogeneous” products are declining in popularity. As one CEO of a reinsurance company stated: “It’s [bouquets] something of the past. There are a few bouquets around Eastern European ... but this is being phased out more. Because bouquets are simply blurring the picture, you want to ensure that you look at each business on its own merits ... I don’t think there’s a future for bouquet placements at the moment” [Reinsurer]. While, they remain in demand in some emerging markets and from some small insurers, where firms have risks that are too small to be attractive to reinsurers on a standalone basis, they remain a small and decreasing part of the reinsurance market.

Stand-alone products are also a traditional product that used to dominate reinsurance. However, these single line and territory products are under threat as local market buyers decline and larger players seek greater efficiency in their reinsurance spend through centrally coordinated purchasing. Specifically, as global players acquire local companies, standalone reinsurance products disappear from the market, and are being replaced by their new parents’ bundled “super-risks”, which we now discuss.

Implication 2: Increasing global and regional covers (super-risks)

The dominant players in the industry – large regional and global insurance players – require bundled homogenous products that we term “super-risks”. These super-risks are popular with large insurers because bundling of homogeneous risks, such as catastrophe risk across perils and territories, into single “super-risks” products, enables capital efficiency and increases reinsurance purchasing efficiency through greater coordination. For instance, the resurgent American International Group (AIG) is undertaking a significant consolidation of its property per risk reinsurance by buying a single U.S.$1.5b global treaty, further demonstrating the trend of major insurers to streamline their own risk transfer into single super-risk products.¹⁵

Super-risk products are, therefore, growing, indicating the effects of consolidation in the primary insurance industry upon the reinsurance product. These products then replace the standalone local products (outlined above) of local companies that have been acquired. Namely, as they acquire new companies, reinsurers consolidate all the separate reinsurance programs that these smaller insurers used to buy, as one CEO of a major global insurance company explains: “So in the last 10 years [we] have doubled in size with acquisitions in the States and Europe. Our original model was probably lots of small businesses that all bought their own programs. But as we’ve been through this expansion program, virtually every acquisition we do, you get a reinsurance synergy from either not having to buy the program or combining it with your existing program. So, there’s been an evolution toward combining programs through our acquisitions. A few years ago we realized that allowing our businesses to pick and choose was inherently inefficient. So we decided to set out and do a global deal” [Cedent, interview]. Importantly, from a reinsurance perspective, as global reinsurers shift to centralized buying it means that premiums no longer flow from local programs and the overall amount of premium ceded is reduced, as one reinsurer explains: “They went from buying proportional to non-proportional and stopped ceding treaties from local operating entities. A lot of meat from the bone was gone for us; we lost 50% of our premium just due to this” [Reinsurer, interview]. Such dynamics help explain the increased competitive pressures on reinsurers outlined in the first section of this article.

Summary

Overall, our research suggests the following trends in terms of cedents’ reinsurance product needs: 1) as some of the “traditional” reinsurance products mature, such as bouquet (single territories, multiple types of risk) and standalone (single risk-type single or similar territories) products, demand will progressively decline; 2) the demand for other products, such as super-risks (covering multiple, or even global territories) will grow. Such changes require reinsurers to keep abreast of change in order to match their supply of capital with changing patterns of demand. We, therefore, end with further implications of such trends for reinsurance executives.

The Journal of Financial Perspectives

Final considerations on reinsurance product evolution

There are three main issues that arise from the super-risk product and super-cedent growth, which we highlight below. We also suggest ways in which reinsurers can reflect upon these issues in relation to their specific strategic positions and goals.

Issue 1: Pressure on smaller reinsurers

From the reinsurer’s point of view, there is a shift from multiple standalone products each yielding smaller, territory-specific premiums, to fewer, larger multi-territory reinsurance products each yielding large premiums. The distribution of global premium thus also shifts toward these super-risks – forcing more and more reinsurers to find ways to offer such products. This puts pressure on smaller players in the reinsurance industry, who struggle for relevance because they can only write a tiny fraction of the overall deal. They, therefore, may be ignored, or squeezed out of such deals. This difference in perspective is outlined from the perspective of a large cedent and a small reinsurer:

• Large cedent: “We felt we had such long-term strong relationships with three or four global players that we should leverage to cement a proper long-term relationship and make them our core providers. So, we’re really a believer in supporting the stronger balance sheets, the quality players,” (Cedent interview).

• Small reinsurer: “If we talk about the 10 biggest insurers; if you look at their programs it’s going to be a worldwide Cat XL: $2 billion of capacity. What can we do? How can we be with them? We just haven’t got the modelling and also capital and capacity to be a meaningful partner to them.” (Cedent interview).

For example, the majority of QBE’s global programs are placed with three large reinsurers; and recently, much of the remaining available premium (20%) has been placed with Berkshire Hathaway.17 In short, the above changes may be responded to more effectively by larger reinsurers, while other reinsurers may struggle to maintain meaningful partnerships with such large cedents. Nonetheless, as Table 1 shows, smaller reinsurers may still be well positioned to exploit niches in fast-growing emerging markets through in-depth client knowledge.

Issue 2: Change in reinsurance operations and resources

Super-risks are a complex product, covering multiple territories and perils. This has interrupted the direct link between the origin of risk and its cover that is prevalent in stand-alone products. It has also made it more difficult for the requisite knowledge to underwrite a deal to reside in a particular person or role (i.e., a single underwriter). Instead, reinsurers have to increasingly

---


---

<table>
<thead>
<tr>
<th>Cedent type</th>
<th>Product type</th>
<th>Meaningful partnerships</th>
<th>Implications for different types of reinsurers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global [growth]</td>
<td>Super-risks and ART</td>
<td>Large - highly analytic - reinsurers</td>
<td>Highly analytic; Large player</td>
</tr>
<tr>
<td></td>
<td>Super-Risks and ART</td>
<td>Medium/Large - highly analytic - reinsurers</td>
<td>Comparatively comfortable; Larger and – to a lesser degree – moderate sized reinsurers</td>
</tr>
<tr>
<td>Regional [growth]</td>
<td>Local programmes</td>
<td>Meaningful partnerships possible with all types of reinsurer</td>
<td>Difficult space: smaller/less analytic reinsurers who relied on these</td>
</tr>
<tr>
<td></td>
<td>Bouquets</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Pushed out increasing demand for analysability]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local [Decline]</td>
<td>Local</td>
<td>Less analytic (close to market) reinsurers</td>
<td>Low-tech; size less relevant.</td>
</tr>
<tr>
<td></td>
<td>Bouquet</td>
<td></td>
<td>Niches still available through local market knowledge</td>
</tr>
<tr>
<td>Emerging [Small in premium terms]</td>
<td>Local</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bouquet</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Matching reinsurers and insurers: who is most impacted by trend
work in teams and in particular rely on technical resources to analyze these complex deals. Super-risks are grounded in more complex financial modeling than existing products, and so require different competencies and technical infrastructure from the reinsurers underwriting them than standalone products. As super-risk programs increase, fewer reinsurers can handle the level of analysis required. Reflecting on the analytical competences that comes with the complexity of such deals, one manager of an insurance firm said of his super-risk deal: “Getting the risk home was more difficult just because it’s hugely complex and if you like it, you know, the [Company 1 and 2] put six months’ work in to it to understanding it” [Cedent interview]. If reinsurers are to reposition themselves to offer super-risk products, they need to acquire these skills and competencies and change their operating structure accordingly.

Issue 3: Growth of ART products that compete with reinsurance

ART products are usually thought as “non-traditional” reinsurance products in that they are not generally offered by reinsurers, but instead by hedge funds and other providers of capital investment. Yet, these products are in a “growth” phase in the industry [Aon Benfield (2013a)] and they are no longer separate from the reinsurance. For instance, they are an attractive option for larger cedents for a number of reasons, including diversifying between traditional and non-traditional risk carriers. For example, as the CEO of Allianz Re explains: “About one tenth of our overall natural catastrophe protection comes from Cat bonds. In conjunction with our protection against US natural catastrophe risks, the share of Cat bonds is considerably higher, about one-third.” ART products are securities whose returns depend on the occurrence of a specific insurance event [Fedorova (2012)]. This is often what large insurers require capital for: peak risks which could significantly impact their profitability or even survival. They are also well suited to the complex reinsurance, technicality and capital efficiency needs of larger cedents.

In short, from a reinsurer’s point of view, ART products compete with traditional reinsurance products; and ILS products (which include Cat bonds) are partial substitutes for traditional reinsurance. These partial substitutes, which are likely to continue to grow in their market share, alter the reinsurance product life cycle in ways that threaten existing standalone and bouquet products: as more premium goes into super-risks and ART, a declining share is left for standalone and bouquet products. It is clear that these trends have different implications for reinsurers depending on their positioning; for instance, they are particularly worrisome for late adopters and those without the technical or capital resources to participate fully in such a suite of products [Aon Benfield (2013b)].

Summary: Recommendations for reinsurers

The above discussion implies a number of recommendations for reinsurers, which we summarize here:

- As available premium from traditional “standalone” products decreases, reinsurers will need to protect their existing relationships with their most desirable local clients. They will, however, also simultaneously need to establish alternative bases for generating the volume of their revenue through developing the resources to partner with larger global clients.
- Smaller reinsurers need to be particularly mindful of defining a profitable niche for themselves in response to these trends. Given that they might be less obvious partners for large global firms, they need to look to niche areas where their size means they can offer bespoke services or specialize in particular areas of expertise. For instance, in a particular emerging market or specialty line of business. Smaller reinsurers might also look for appropriate joint ventures and partnerships with other capital providers in conjunction with developing and protecting profitable niche areas.
- To analyze mega-risks and meet the needs of global clients, it will be necessary for reinsurers to continue to invest heavily in analytical infrastructure and expertise. Those that do so will be the preferred partners of such global insurers because they are able to address their more complex needs.
- Reinsurers will increasingly be required to engage with ART products and new players (e.g., hedge funds) entering the reinsurance space. Those that are already proving more innovative in their approach to capital provision and cover will

---

be better placed in the coming years. This includes seeking the necessary expertise, partnerships and structures to position themselves as potential firms to front the delivery of these newer products.

**Conclusion**

In this article, we have first highlighted the general trends resulting in increasing competitiveness in the reinsurance industry; using a strategy model [Porter (1980)] to structure our discussion. Second, we focused on a particularly salient area from which much of these competitive dynamics stem: consolidation in the primary industry, which we describe as part of life cycle analysis of the insurance sector. Third, we highlighted how the industry life cycle in the insurance sector is impacting reinsurance buying, thus having important impact on the reinsurance industry that reinsurance executives should be particularly mindful of. Finally, we outlined some specific implications for reinsurance executives and firms, based on this analysis of changes in the product life cycle.

Through building an overview of the industry and focusing on consolidation in the primary sector, we sought to put some of the strategic decisions and difficulties facing reinsurers in context. This should provide those reinsurers with greater understanding of the challenges they face and, thus, how to cope with them.

**References**


Aon Benfield, 2013b. “Increasing velocity of new capital flows benefits reinsurance buyers,” reinsurance market Outlook, Aon Benfield


In search of a more stable monetary and financial order

Thomas Mayer
Senior Advisor, Deutsche Bank and Senior Fellow, Center of Financial Studies, Goethe University Frankfurt am Main

Abstract
Policymakers in governments and central banks and mainstream economists want to improve the stability of our monetary and financial system by more public intervention in, and regulation of, private financial markets. Implicit in this approach is the premise that too little intervention and regulation caused the financial crisis. This paper argues that the opposite is true: it was too much and misguided intervention that caused the crisis. Hence, to make our monetary and financial system safer we need better frameworks for monetary policy and bank regulation. In a new framework for monetary policy credit must be given a prominent role, and in a new framework for regulation we must properly define safe and loss-absorbing assets and use the latter to cover the costs of bank failures (instead of passing the bill to the taxpayers). When the costs of failure are internalized, banks will be under pressure to become more transparent for creditors and depositors and their cost of funds and lending rates will rise. But higher costs of funds for banks will only reflect the end of the public subsidy banks benefit from, as a result of the possibility of taxpayer funded bailouts.
1. Introduction
A popular view is that “light” regulation of the financial industry during the 1990s and early 2000s allowed the financial excesses and rise in indebtedness that eventually caused the financial crisis. Those who hold this view want regulation to be tightened to prevent another crisis in the future. Thus, in 2008, the G20 demanded that no actor, no product and no market remain unregulated. Since then, authorities have been active at the national and international level to fulfill this promise. However, there is good reason to believe that the premise for more regulation is wrong. In this paper, I shall argue that the popular diagnosis of the causes of the financial crisis is flawed and hence the suggested therapy misguided. In my view, it was not too little but too much official intervention in financial markets that caused the crisis. Instead of more public meddling in private markets, we need a stable framework that allows these markets to function properly.

To lay out my argument, I shall discuss the nature of fractional reserve banking and the need for government regulation and intervention within this system in the following section. In section three, I review the approach to monetary and financial sector policy in the run-up to the financial crisis. In section four, I discuss the efforts to stabilize the economies hit by the crisis, and in section five I present a few ideas for a better monetary and financial system. Section six concludes the paper.

2. The need for regulation of fractional reserve banking
Fractional reserve banking has been the standard business model of the banking industry since the Middle Ages [Huerta de Soto (2009)]. Originally, bankers took deposits in the form of metal coins and issued certificates, and later paper bank notes, against them. But instead of keeping all deposited coins in the bank, they kept only a fraction and lent out the rest, hoping that the depositors would not want to exchange book money into central bank money than banks can accommodate with the reserves of central bank money they hold. In this case, banks, or the entire banking system, may face a liquidity crisis. Second, credit that banks have extended to create book money may not be repaid. When the write-off of credits exceeds the amount of equity capital banks hold as a buffer, they, or the entire banking system, may face a solvency crisis. In the course of history, central banks were created to deal with liquidity crises, and publicly managed deposit insurance schemes were built to deal with solvency crises. Thus, in 1654, the King of Sweden took over a private bank, which had issued bank notes against deposits of copper coins and experienced a liquidity crisis. He turned it into the first central bank, later named the Swedish Riksbank, by declaring the notes issued against the copper coins to be legal tender. Centuries later, against the background of many bank failures during the Great Depression, the Roosevelt administration in 1933 created the Federal Deposit Insurance Corporation to protect depositors against the consequences of insolvency of banks. Although members of the so-called free banking school do not want to accept this [See, for instance, Selgin (1988)], history suggests that in a fractional reserve banking system official backstops through central banks and governments are needed to protect the general public from the fall-out of recurrent liquidity and solvency crises.

Today, credit and money are created in a public-private partnership. By setting interest rates on their loans of central bank money to commercial banks, central banks influence bank lending rates and, through this credit, demand by the non-financial sector. In satisfying this demand, commercial banks create book money. In times of liquidity or solvency crises, central banks and

---

1 The nature of legal tender is that it must be accepted by creditors for the settlement of a debt, even though it has no intrinsic value of itself. For instance, if I owe a worker an hour’s wage I do not have to give him food or another equivalent commodity, but can give him a paper note with only a nominal value that has been declared legal tender by the government.
governments stand ready to provide liquidity support and protect depositors. Given their close relation with commercial banks and the need to provide potentially large financial support in times of crises, it is no surprise that governments and central banks have a keen interest in the business practices of commercial banks and want to steer banks’ behavior through regulation.

3. The policies that led to the crisis
In the course of the 1980s, a new economic paradigm emerged that combined elements of the older Keynesian and neoclassical theories with the newly developed theory of rational expectations. Central banks, which had in the early 1980s flirted with market-liberal monetarist theories developed primarily by Milton Friedman, adopted this paradigm – let us call it the New Keynesian-Neoclassical Synthesis (NNS) – in the course of the late 1980s and 1990s. According to Goodfriend and King (1997), the NNS is based on four elements: (i) inter-temporal optimization; (ii) rational expectations; (iii) imperfect competition; and (iv) costs of price adjustments (menu costs). It is easy to see that the first two elements are inspired by neoclassical theory while the last two are imported from Keynesian theory. In a later description of the NNS, Woodford (2009) pointed out that (i) intertemporal optimization combines both short- and long-term as well as microeconomic and macroeconomic elements in one approach; (ii) theoretical models are empirically calibrated; (iii) the assumption of rational expectations overcomes the Lucas critique; (iv) price and other rigidities allow short-term real effects of a change in money supply; (v) business cycles result from a variety of shocks, not limited to one real or monetary shock; and (vi) the central bank can control inflation, which is explained in an augmented Phillips curve model. For the purpose of interest rate policy in an inflation targeting approach to monetary policy, a simple rule can be derived, in which the policy rate is determined by a cyclically neutral real rate, the inflation target, the output gap and deviations of actual inflation from target inflation [Taylor (1993)].

The assumptions of rational expectations, efficiency of financial markets, and neutrality of money in the longterm allow the NNS to do without a banking sector and a capital market. Money is simply the catalyst for real economy transactions with no (long-term) effects of its own, the proverbial “veil” over the face of the real economy. Banks and capital markets serve savers and investors by connecting them to each other without playing an active role in the economy. In other words, money, banks and capital markets lubricate the real economy to make it work better, but do not influence it as actors of their own. Hence, central banks basing their monetary policy on the NNS in an inflation targeting framework are blind to developments in banking and finance that are inconsistent with the passive role ascribed to them by theory. Moreover, since financial markets are presumed to be efficient and economic agents to have rational expectations – and therefore know better than regulators what to do – there seems to be no need for heavy regulation of the financial sector. All that seems necessary is that bankers and financial market participants observe some minimum standards of business conduct.

In the course of the 1990s, technical progress and the integration of major emerging market economies, notably China and India, into the global economy led to downward pressure on consumer prices. Like in the 19th century, during the process of industrialization, the deflationary pressures originated from the supply side and hence were positive in nature, in contrast to the demand-side induced deflation as a result of “debt deflation” during the Great Depression. However, the major central banks in the industrial countries reacted to the downward pressure on prices by lowering interest rates to achieve their positive inflation targets. They did so by mainly inflating prices of non-tradable goods. In the process, they induced a rise in asset prices and stimulated credit demand. Thanks to the ample availability of cheap credit, occasional bumps in economic activity due to the volatility in asset prices could be quickly overcome.

The combination of an activist stabilization policy for the real economy and a laissez-faire attitude toward the financial sector, which is embedded in the compromise between the New Keynesians and Neoclassics that gave birth to the NNS, created moral hazard. The more successfully that the business cycle could be smoothed by monetary policy, the less risky rising leverage in the financial sector became. If in addition monetary policy was prepared to specifically bail out leveraged investors that made wrong financial bets, the appetite for leverage and risk would surge to hitherto unknown levels. Former FOMC Chairman Alan Greenspan initiated a monetary policy approach that provided insurance to leveraged investors, who returned the favor by calling this the “Greenspan Put.” His successor, FOMC Chairman Ben Bernanke, provided the scientific foundation for the “Greenspan Put.” In a 2001 paper, he and his co-author Mark
Gertler employed a dynamic stochastic general equilibrium model with New Keynesian features and concluded “that for plausible parameter values the central bank should not respond to asset prices,” but concentrate on stabilizing the real economy through inflation targeting [Bernanke and Gertler (2001)].

4. More of the same
The combination of interventionist policies in the real economy and laissez-fair policies in the financial sector paved the way for what was probably the greatest global credit bubble in history. When the bubble began to burst in the sub-prime segment of the U.S. mortgage market authorities were quick to point out that it was the fault of insufficient regulation and control of greedy bankers. Supported by mainstream economists relying on the NNS as their model for the economy, they set out to extend the interventionist approach of the NNS from the real economy to the financial sector. Measures to better regulate the financial industry include, among others, (i) higher requirements for equity capital of banks; (ii) the need for banks to hold more liquid assets; (iii) the separation of trading activities from deposit taking and credit extension; (iv) caps on bankers’ bonus payments; (v) better consumer protection through detailed documentation of expert advice; (vi) participation of the banking sector in the costs of financial crises; (vii) a resolution regime for banks; and (viii) product regulations to increase transparency and better align the interests of issuers with those of investors. Although the work is not finished, one may conclude already that the final product will suffer from excessive complexity and inherent contradictions.

Moreover, as commercial banks experienced heavy losses on credit extended during the upswing of the credit cycle they saw their equity cushions shrink and, in some cases, disappear. Governments stepped in to recapitalize banks. With equity capital down and equity requirements raised to higher levels than before the crisis, banks cut down on credit extension to shrink their balance sheets and raise equity ratios to the ambitious levels demanded by regulators. As credit to the non-financial sector began to dry up, central banks stepped in and provided credit to the public and private sector directly (the latter to a large extent in the form of mortgage credit). The balance sheet extension of central banks was accompanied by a reduction of policy rates to record low levels and promises to keep rates at these levels for an extended period of time (or until certain economic indicators, such as unemployment rates, reached desired targets). Thus, central banks joined regulators to create a largely centrally planned global financial system.

To sum up, policy interventions and misguided focus of the regulators created moral hazard and were key ingredients in the emergence of credit excesses that resulted in the financial crisis. However, instead of backtracking and leaving it to markets to reward success with profits and punish failure with losses and bankruptcies, authorities have set out to extend the interventionism reserved for the real economy in the past to the financial industry. With this, however, they will augment, rather than correct, their past mistake of presuming that they are central planners in possession of knowledge they cannot possibly have. Future generations of economists may find it ironic that after the failure of centrally planning the real economy in socialism a little less than three decades ago, the capitalist countries have set out to establish central planning in the financial sector. Like real economy central planning in socialism in the past, financial sector central planning in capitalism now must fail due to the inferior knowledge of the central planners.

5. A more robust financial system for the future
To make our financial system more robust and avoid another credit cycle we need significant changes in the approach to monetary policy and to regulation. The former needs to become more modest and focus on variables it can influence rather than to aim for real economy goals, such as economic growth or unemployment, over which it has little control. The latter needs to abstain from interfering in the daily business decisions of financial institutions and, instead, must concentrate on establishing key principles for the conduct of business in the financial industry.

The main reason for the rise of credit and debt to unsustainable levels was that central banks ignored credit developments even though they exerted a strong influence over them. Hence, central banks should either get out of the way entirely or bring credit into the focus of their policy. It is not the purpose of this paper to discuss proposals to abolish central banks as they are unlikely to be implemented irrespective of their merit. Central banks have evolved over hundreds of years because they serve the interests of governments and banks and are unlikely to be disbanded under any circumstances short of a revolution of our economic and financial system. Reforms must, therefore, concentrate on giving credit the prominent role in the design of monetary policy that it deserves.
In a recent paper, Michael Biggs and I have argued that contrary to conventional wisdom the flow of credit is highly correlated with demand [Biggs and Mayer (2013)]. This holds across many countries and long periods of time. Mainstream economists have overlooked this relationship because (i) they were blinded by their models, in which credit plays no active role in the economy; and (ii) they erroneously focused on the relationship between credit stocks (instead of flows) and demand flows, which seemed to support the passive role ascribed to credit by their models. We also showed that while credit flows are correlated with demand flows the growth of the credit stock is correlated with the output gap.

The respective variables are shown in Figures 1 and 2 for the U.S. Figure 1 relates the growth of real domestic demand to the change in credit flows (normalized by GDP), a measure we call “credit impulse” because of its active influence on demand. The Figure shows how big variations in the flow of credit went along with strong variations of real demand during the recent financial crisis as well as during the Great Depression of the 1930s. Figure 2 relates the output gap to the growth of the credit stock. From this Figure it can be seen that an output gap of zero, the objective of central banks following the approach of inflation targeting, was accompanied by credit growth of 7% during the last three decades. Since nominal GDP growth averaged 5% during this period, a monetary policy that succeeded in maintaining full and stable capacity utilization allowed the ratio of credit to GDP to rise without limit. Sooner or later this policy had to end in a financial crisis caused by excessive credit growth.

Against this background, we suggest to bring credit into the focus of monetary policy. In our view, central banks should aim for steady credit growth in line with a sustainable ratio of credit to GDP. The latter is, of course, subject to judgment, but judgment can be formed on the basis of historical experience (i.e., by looking at periods where the ratio of credit to GDP went along with stable developments of the real economy and the financial sector). If the ratio of credit to GDP is low at the starting point, monetary policy can aim to stabilize credit growth for a while at a rate above its long-term sustainable rate until it has reached the latter. If the ratio of credit to GDP is high at the starting point, the opposite applies. By adjusting interest rates and other tools in response to credit developments (instead of in response to developments of the real economy), monetary policy remains in the domain where its influence is most pronounced while contributing to economic stability by aiming for price and financial stability.

The new approach to monetary policy needs to be complemented by a regulatory framework that focuses on closely linking risk taking to return in the financial sector. This requires first and foremost the possibility of financial institutions...
In search of a more stable monetary and financial order

To fail without pushing the entire economy into recession or even depression. As I have argued in a recent paper, the starting point for such a regulatory framework is the definition of a really safe asset [Mayer (2013)]. An asset is really safe when it can be converted into legal tender at face value any time and under any circumstances. Only an asset with these qualities ensures that debts can be settled when they are due. In our system of artificially created (fiat) money, only central banks have the right to issue legal tender and hence only assets that can always be turned into central bank money are really safe. From this follows that in a regime where governments, like private entities, have no access to the money printing press to settle their debt, only cash and bank deposits fully covered by reserves held with the central bank are safe. Thus, when we define government bonds as safe assets, as is generally done in finance textbooks, we assume that the government can command the production of legal tender by the central bank to secure the exchange of these bonds into legal tender when needed. Hence, it is inconsistent to legally ban the access of the government to central bank credit and, at the same time, to define government bonds as safe assets.

When we have defined the safe asset for the general public it is easy to establish a hierarchy of loss-absorbing bank liabilities in a bank resolution regime. Clearly, the first loss must be covered by bank equity capital. When that is wiped out further losses are covered, in consecutive order, by junior and senior unsecured debt and, finally, by deposits not fully covered by central bank reserves (let us call them “investor deposits” in contrast to the fully covered “safe deposits”). Table 1 gives a stylized balance sheet of a bank under this regime.

Money in a safe deposit is, like cash, in fact a liability of the central bank. Banks only act as “safe-keepers” of funds held in safe deposits. Hence, only the central bank – and not commercial banks – can create safe deposits. An initial endowment of safe deposits can be created by the central bank through purchases of low-risk assets, notably government bonds, from commercial banks. In the transition from the present to the new regime purchases of government bonds by the central bank would help commercial banks reduce their holdings of government bonds to levels consistent with the nature of these bonds as risky assets, i.e., assets with a non-zero probability of default.

Table 1: Stylized bank balance sheet
Source: Mayer (2013)

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central bank reserves</td>
<td>Safe deposits</td>
</tr>
<tr>
<td>Ring-fenced assets</td>
<td>Covered bonds</td>
</tr>
<tr>
<td>Other assets</td>
<td>Investor deposits*</td>
</tr>
<tr>
<td></td>
<td>Senior debt*</td>
</tr>
<tr>
<td></td>
<td>Junior debt*</td>
</tr>
<tr>
<td></td>
<td>Equity*</td>
</tr>
<tr>
<td></td>
<td>* Participating in losses in ascending order</td>
</tr>
</tbody>
</table>

Table 2: Central bank balance sheet
Source: Mayer (2013)

<table>
<thead>
<tr>
<th>Assets</th>
<th>Liabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government bonds</td>
<td>Deposits of commercial bank reserves to cover safe deposits</td>
</tr>
<tr>
<td>Good will</td>
<td>Other liabilities, reserves and capital</td>
</tr>
<tr>
<td>Other assets</td>
<td></td>
</tr>
</tbody>
</table>

Safe deposits can be increased further when commercial banks borrow reserves from the central bank to create safe deposits. Since commercial banks borrow reserves to deposit them immediately with the central bank, they need not post collateral to obtain the loans. But the creation of safe deposits through borrowing of reserves comes at a cost, given by the difference between the central bank’s lending and deposit rates for central bank money. Banks would, of course, pass these costs on to their customers demanding safe deposits. Thus, by varying the difference between its lending and deposit rate the central bank can influence the demand for safe deposits. It can narrow the spread and make safe deposits cheap during an economic upswing when the general public is willing to take risks and puts a low price on safety. And it can widen the spread and make safe deposits expensive during a downswing when the general public puts a high price on safety. By influencing the mix between investor and safe deposits through interest rate policy the central bank can also affect the credit and money multipliers and thereby, dampen the credit cycle.

Finally, safe deposits could be increased to satisfy the demand for money in a growing economy by simply paying a dividend to depositors. The book entry on the asset side of the central
bank’s balance sheet for safe deposits created this way would be Good Will, reflecting the trust of the general public in the value of central bank money as a means of exchange and store of value. Table 2 shows the stylized balance sheet of the central bank under this regime.

The new regime requires a much higher degree of disclosure of riskiness of banks’ assets than the present regime, where virtually all bank creditors assume that in the case of bank failures, they are protected by the government or the central bank. In the new regime all bank creditors except the holders of safe deposits have to absorb losses when assets have to be written off. Hence, the funding costs of banks will rise, and with them the cost of credit for non-banks. However, the increase in credit rates is necessary to end the externalization of the costs to the general public of bank failures due to reckless lending. When these costs are internalized, credit will be allocated more efficiently and credit bubbles will become less likely. Internalization of the costs of bank failures will prevent excessive complexity of bank balance sheets, and eliminate banks’ desire to become “too big and too interconnected to fail.”

6. Conclusion
Policymakers in governments and central banks and mainstream economists want to improve the stability of our monetary and financial system by more public intervention in, and regulation of, private financial markets. Implicit in this approach is the premise that too little intervention and regulation caused the financial crisis. In this paper, I have argued that the opposite is true: it was too much and misguided intervention that caused the crisis. Hence, to make our monetary and financial system safer we need better frameworks for monetary policy and bank regulation. In a new framework for monetary policy, credit must be given a prominent role; and in a new framework for regulation we must properly define safe and loss-absorbing assets and use the latter to cover the costs of bank failures (instead of passing the bill to the taxpayers). When the costs of failure are internalized, banks will be under pressure to become more transparent for creditors and depositors, and their cost of funds and lending rates will rise. But higher costs of funds for banks will only reflect the end of the public subsidy that banks benefit from as a result of the possibility of taxpayer funded bailouts.

References
Rethinking the future of financial services: a structural and evolutionary perspective on regulation

Michael G. Jacobides
Sir Donald Gordon Chair of Entrepreneurship and Innovation and Associate Professor of Strategy and Entrepreneurship, London Business School

Michael Drexler
Senior Director and Head of Investor Industries, World Economic Forum

Jason R. Rico
Senior Project Manager, World Economic Forum

Abstract
The fundamental reasons why financial crises start may not have changed much over the centuries, given the immutable qualities and pathologies of human nature. However, the nature of the financial system has changed – it has become secularly more interconnected, immediate, global and complex. The regulatory climate had not adjusted to these changes, leaving us ill-prepared to prevent the recent financial crisis. For all the rhetoric to the contrary, regulators’ frameworks continue to be based on an institutionally outdated understanding, poorly suited to tackle the dynamic nature of the financial ecosystem and its changing roster of participants. Thus, the next crisis may already be brewing; and despite the hostility to banks and bankers, the system is not as resilient as it could be. We argue that the regulatory framework has to be reconsidered, so as to better track and respond to the changing industry architecture of financial services. Our paper provides an appreciative analysis of what happened during the recent crisis, and offers recommendations for a regulatory rethink. We illustrate this by considering the changing architecture of the financial system, paying particular attention to the evolution of new business models over the last decades, both in the regulated part of financial services and the alternative investments ecosystem that has co-evolved with it. We cast the recent financial crisis and regulatory responses in the context of this framework, and outline an ecosystem-based approach that looks at the dynamics of the sector’s architecture. We close with recommendations for the regulatory approach and practice.
Why do we need a structural and evolutionary perspective?
In the immediate aftermath of the financial crisis, a group comprising some of the most respected economists in the U.S., noted that “... the crisis revealed both gaps in regulation and unanticipated interconnections among different types of financial institutions and markets. Yet, no one was charged with understanding these interconnections, looking for gaps, detecting early signs of systemic threats and acting to mitigate them.” [Pew Charitable Trust, Financial Reform Project Task Force (2009)]. Five years later, one would have hoped that both analysis (in terms of economics and finance) and policy might have taken stock of what happened, accepted the lessons, and changed accordingly. Yet, relatively little appears to have changed in the way that we understand the financial system and its regulation, whether on the academic or the policy front [Colander et al. (2011)]. This paper aspires to offer a complement to the mainstream analysis, and identify a set of concrete recommendations for action which, if implemented, might have averted the previous crisis, or might avert the next crisis, which could already be in the making.

We start by offering a systemic view of the financial sector and of the nature of risk. We then move to an analysis of the drivers of the sector’s evolution, focusing in particular on its “industry architecture,” the set of rules and roles that pertain to the division of labor, and which circumscribe the business models in place. We then explore the systemic interactions, before summarizing the current regulatory response and challenges. We conclude with our analysis of the challenges of the regulatory framework, and offer concrete recommendations on how to solve some of these challenges, based on our evolutionary structural approach.

A systemic view of the financial sector
For better or worse, the global financial system pervades nearly all aspects of economic activity. From ledger- and credit-based origins in Mesopotamia and very likely even earlier human societies [Graeber (2011)], through the development of coinage and ever more sophisticated financial instruments to the present day, the financial system has tracked economic development and wealth creation. Independent of its sophistication, three major functions describe its purpose to society: (1) facilitation of economic transactions (e.g., via payments systems); (2) allocation of capital (in particular from savers to borrowers) over time and space; and (3) transfer of risk between economic actors. Of those, the first is often referred to as the system’s “plumbing” and covers a vast majority of money flows every day. It is of obvious importance to a well-functioning global economy. However, this paper focuses on the latter two – not only are they more complicated, but they also involve the majority of systemic issues of the financial system that came to light in the recent crisis.

At a macro level, it is very easy to describe the global financial system, as Figure 1 illustrates. This familiar picture serves to emphasize two points. First, that the financial system is a complex dynamic system with feedback loops between actors. Second, that intermediaries (whether regulated or not) pool capital flows and design products that best suit the needs of other actors – final users of capital, intermediaries and their own. Intermediaries’ positions both mirror the needs of the users and suppliers of capital, but also arguably shape these needs – for example, through the creation of money and the provision of credit for particular activities.

Understanding the nature of these intermediaries, and devising policies and frameworks to deal effectively with them, is central to effective regulation. So far, the practice of regulation has tended to look at “one type of intermediary at a time”, especially in the large economies where the majority of the financial assets reside – namely, the U.S., the U.K. and the E.U. In the U.S., in
In reality, though, this approach seems to ignore the systemic nature of the sector which we just outlined. The last crisis provides a perfect illustration. For all the rhetoric of the risks of integrated banks, the recipients of state support in the U.S. included major non-depository institutions, many not even part of the banking system: Bear Sterns, Fannie Mae, Freddie Mac and AIG were all specialized intermediaries, whose interconnected nature led the Federal government (and the FED) to intervene to support them so as to preserve the viability of the system. Even the collapse of Lehman Brothers was a collapse of a strictly focused wholesale bank, with strong bonus claw-back policies, and no links to retail customers. This suggests that in past crises, and most probably in future crises, it is the systemic nature of the sector’s constantly evolving structure that may lead to instabilities. In turn, regulations must also continuously evolve, lest regulators be left to focus narrowly on a small part of the sector and lose the forest for the trees, as we argue happened in the last crisis. Both regulations and regulators’ skills must keep pace with the changes in the financial sector.

**Risks from a systemic perspective**

Moving from a critical approach to a positive set of recommendations, we consider that financial service disruptions should be considered, at the level of the entire ecosystem, on the basis of risk-inducing shocks. As in any complex system, shocks can occur, and they normally have one or more of the following causes: (1) loss of confidence in a significant share of capital sources (e.g., threat of currency devaluation), destinations (e.g., corporate credit quality) or intermediaries (e.g., banking crisis); with the latter being the most frequent and systemically catastrophic due to the higher concentration of the intermediary landscape; (2) sustained imbalances in capital flows, particularly across different time horizons (contingent claims); and (3) inability of intermediaries to appropriately channel capital flows.

Fundamentally, the system’s top-level intricacy lies in four major issues (several more issues arise at more granular levels):

- A mapping of actors – some institutions can be both sources and destinations of capital, and sometimes engage in intermediary functions also, as shown above.
- Highly adaptive and intelligent (from a perspective of maximizing own returns) actors who often have better human capital than regulatory bodies. In addition, the number of
actors is vast (e.g., 7,000 banks in the U.S. alone) which makes for a highly interconnected and complicated system at a more granular level.

- Rapid feedback loops in a system that can enter self-reinforcing negative cycles (aka pro-cyclicality).
- Inter-temporal nature of capital flows and therefore intermediation (introducing time-related risk, such as from “warehousing” mortgage-backed securities).

In order to be able to respond to these challenges, regulators must be able to have a clear and sophisticated understanding both of the institutional structure of the intermediaries sector (i.e., who does what in the financial services ecosystem), and of their interactions (i.e., how failure in one part of the sector affects the other, or indeed how regulation in one part of the sector affects another).

The challenge has been that, to date, we seem to be very far from either of these two goals. As for a grounded understanding of the nature of industry participants, the first official mapping exercise happened a full two years after the crisis — the architecture of the different players and how they map onto each other [(Pozsar et al. (2010)] only gained currency recently, and coincidentally [Jacobides (2014)]. As for the side-effects of regulation, it is even clearer that little was understood ex-ante. For instance, the push for a “Tobin tax” (i.e. a tax on market transactions) ignored the fact that this was not only detrimental to trading activities, but would also impose huge costs to the banks funding themselves from the wholesale market, and was only shelved when the systemic implications became painfully evident, just prior to its potential implementation.

On the basis of these observations, then, it does appear that a more fine-grained understanding of business models in the world of intermediaries and of the evolutionary dynamics and of the feedback loops in the sector might be of use. In the next section, we will endeavor to provide a bit of both, drawing on the evolution of the financial system over the last few decades. Our focus will be the changing business models and dynamics and we will explain how these led to the financial crisis.

Drivers shaping the evolution of the financial system
Beginning in the 1970s, the very nature of the financial system underpinning the global economy began to materially change, giving rise to new structures as well as instruments. No single factor proved decisive. Rather, a confluence of factors such as macro trends in the global economy, regulatory changes across the world, technological innovation, and theoretical breakthroughs in academia began to coalesce. The result was the gradual transition from the historically simple loop connecting savers and borrowers to the highly complex, interdependent, interconnected, and dynamic system that we see today. The new system fundamentally altered the landscape and gave rise to a new set of actors, many frequently misunderstood.

Breakthroughs in academia played a key role in changing the nature of finance and enabled the creation of a wide range of new financial instruments. Financial theoreticians would lay the foundations, with their peers in computer science providing the means to apply the theories in financial markets. Two financial formulas, in the form of the Black-Scholes option pricing model and the copula theorems, would give rise to rapid growth in derivatives [Bernstein (1992)] and structured products [Jones (2009) and Salmon (2009)], respectively. Supported by rapid increase in computational power, the two formulas enabled the financial industry to create new products intended to better carve up and distribute risk among various actors.

Macro-economy driven demands played a key role deepening and expanding the demand for these new instruments – for example, the use of commodity driven derivatives would increase dramatically during the 1970s in response to price volatility of critical industrial inputs. In turn, the early 1980s with its inflationary interest rate volatility saw the emergence at scale of interest rate swaps. The use of credit default swaps would grow rapidly after the dotcom crash and related record corporate default rates. Investment banks played a key role in developing these products in a quest to earn fees in an ever more competitive landscape. Those innovations coordinated hitherto disparate economic actors, creating increasingly interconnected, homogeneous and sizeable markets [Tett (2009)].

Legislative and regulatory actions taken by governments around the world would also prove critical in reshaping the structure of the financial sector. Regulation Q, for instance, enabled the growth of money market funds, which would quickly become a systemic player in the financial system. The collapse
of the Bretton Woods system and the subsequent growth of the Eurobond market in the 1980s, coupled with increasing globalization, favored Universal Banks, a model that started in Europe, and then moved to Japan and finally to the U.S., with the 1999 repeal of the 1933 Glass-Steagall Act by the Financial Modernization Bill (Gramm-Leach-Bliley Act). The law, which was in response to the belief that U.S. banks were losing their competitiveness relative to the European and Japanese peers, enabled the creation of hybrid financial institutions that could combine retail banking, wholesale banking, insurance, and asset management. Critically, it did not require that each line of business be overseen by the regulator historically affiliated with the space. The effect was that wholesale banks in the U.S. could expand into the traditional retail banking space without also having to remain accountable to the FDIC or the Federal Reserve. An era of deregulation in the 1990s and 2000s led to substantial innovation, and to an aversion of looking into the dynamics of financial markets. It also led to a sector that was not only too big to fail but, as some are increasingly arguing, just “too big” [Acharya et al. (2013)].

The growth of derivatives and structured investment vehicles can also be related to government action. The 1999 Financial Modernization Bill exempted derivatives from being regulated, which enabled them to grow largely unchecked by regulatory authorities. Basel regulations also gave a large reduction to capital reserves against securitized assets, when compared with the identical full loans on a banks’ balance sheet; and thus banks found it beneficial to seek greater securitization – which was all too happily provided by the alternative investment or wholesale bank actors active in the booming market [Jacobides (2014)]. Underneath these changes in the financial market, lie individuals and companies intent to reshape the way money was made – shaping new business models to profit themselves, and thus (often unwittingly) reshaping the architecture of the financial system.

In terms of regulatory approach, individually, each new action appeared to make sense, but the growing complexity of the financial system increasingly often resulted in secondary effects far beyond the original intention of policymakers [Pew Charitable Trust, Financial Reform Project Task Force (2009), and Rajan (2011)]. Over time, the gap between how regulators were positioned to govern the system and how the system was actually structured grew, resulting in a significant erosion of the oversight capacity of regulators. This point is illustrated by the growth of alternative (shadow) finance, which until recently had been almost no direct regulation and was thus becoming of potentially systemic relevance as evidenced in the 1998 Long Term Capital Management episode.
Rethinking the future of financial services: a structural and evolutionary perspective on regulation

The systemic shift in the financial ecosystem's architecture

The regulatory, academic and technological shifts fundamentally altered the structure of the financial system, dramatically increasing its complexity and interconnectedness. Historically, the banking system was relatively linear and easy to understand. It consisted of retail and investment banks connecting individual and institutional savers to their matching borrowers. The major role of intermediaries was to monitor the capability of borrowers to pay back their liabilities and reserve adequately for defaults.

However, with the changes outlined in the previous section, as Mehrling et al (2103) find, the system itself has evolved from one focused on funding liquidity to one centered on market liquidity, and one from premised on a “promise to pay” to a “promise to buy.” The resulting credit intermediation process has also grown to include more steps and players, as shown in Figure 4.

The result has been a dramatic increase in the complexity and interconnectedness of the financial system and the steady rise of a vast and largely unregulated shadow banking system that it is interwoven into the fabric of the financial ecosystem (Pozsar et al. (2010)). Figure 5 provides a stylized view of the new ecosystem's complexity; it may be worth noting its original (more legible) size is a 3 by 5 feet. Clearly, the endogenous dynamics of industry change, and the cumulative set of regulatory responses have created an ecosystem of remarkable complexity, albeit not one that is as safe as we would want it to be.

The decrease in the traditional forms of intermediation and commensurate rise of alternative forms of financing can be seen clearly in Figure 6. More important yet, much of the growth in the shadow banking sector has been fed by banks. Banks have originated loans and then used shadow banks to help them optimize their balance sheet [Bord & Santos (2012)], and bank holding companies have seen the majority of the value of their assets be dedicated to activities outside their traditional core [Cetorelli and Goldberg (2012)] — while in the 1990's the vast majority of bank revenues were reflected on their book assets, fee income (and adjusted assets) made up for most of their profitability. Within bank holding companies in particular, the growth came primarily through an explosion of activities in their non-bank subsidiaries, which are evident in the growth of the assets they held. As shown in Figure 7, revenues were reported to be even more skewed toward these new activities. Yet, while universal banks changed their business model and the credit intermediation process evolved dramatically, many regulators maintained frameworks that were anchored in a traditional view of banking. And, ironically, regulation intended to create a safer and more insular financial system often...
Figure 5: Detailed conceptual map of the shadow banking system
Source: Pozsar et al. (2010)
For a full-sized version of this figure, please contact the original source.
unwittingly increased the role of intermediation – both in the regulated financial sector participants and by helping spur the growth of the shadow banking and alternative investment ecosystem [Olson (2012)].

So while the underlying societal needs for finance have remained largely unchanged, the ecosystem addressing those needs has evolved significantly in recent decades. The fact that these changes have also led to a remarkable growth of total indebtedness (i.e., public sector plus individual plus corporate debt) in the countries with the most developed financial systems, is not a coincidence – and suggests that the innovations in intermediation have enabled, in a secular fashion, both the extent of credit (and money supply) available, and the way this is distributed across actors and in time.

A focus on this changing distribution of “rules and roles” in the sector, in the “stable but evolving rules that pertain to the division of labor” that was dubbed “industry architecture” by Jacobides, et al. (2006), merits our attention. This was a case of a dynamic system evolving as a result of endogenous pressures, with the forces of change coming from within. Innovators, seeking profit and margins, drove change; regulators, motivated by politics and political or academic beliefs (that “the market” is superior) allowed for unsupervised innovation; and new structures took hold. Traditional actors adjusted their business models, new sets of actors and related products emerged – all actively seeking to maximize their returns within (or around) the regulatory guidelines in place. Not surprisingly, there has been a commensurate increase in the number and complexity of interconnections within the financial ecosystem.

This brings us to the changing structure of the world of non-traditional finance and alternative investments – in 1980, the channel barely existed, but 30 years later it has grown to provide more than U.S.$6t worth of capital. In order to understand those changes, not only do we have to consider the incentives and evolutionary trajectory of the regulated part of financial services, which we reviewed earlier, but also consider the global macro-economic shifts that have shaped the financial sector, which explain the appetite for products traded through the global capital markets.

Political and institutional changes enabled many non-OECD countries to enter the global economy for the first time, leading to a dramatic increase in the world’s effective labor supply. Export oriented strategies and off-shoring were the natural result, with goods produced in emerging markets being predominantly sold to citizens in OECD nations. Also, as the West was reducing its savings rates, the newly developed East (especially China) was accumulating capital in search of investment options outside the
The resultant savings and trade imbalances left emerging markets with large amounts of U.S. dollar and euro, which were subsequently loaned back into the OECD through sovereign debt and other vehicles. The cycle was self-reinforcing, as each turn pushed interest rates lower and encouraged the further accumulation of debt, much of which was used to purchase more imported goods from emerging nations. On the financial side, this facilitated the vertical disintegration of financial institutions and the reliance of banks on wholesale funding. New instruments, new markets and new business models, both within and outside the regulated part of financial services resulted. The regulatory, academic and technological changes within the financial sector reinforced the trend. Total indebtedness in Western countries grew, and nobody wanted to spoil the party.

The banking, asset management and insurance sectors were impacted significantly. Universal banks sought to provide a global one-stop shop to their clients, while benefiting from diversification and funding benefits that were believed to come with this scale. And, as the alternative investment sector grew, banks sought to emulate compensation practices to retain talent, increasing pay and making incentives stronger than they had been before, potentially changing the behavior within the regulated parts of the financial services sector (Jacobides (2014)).

These changes also were visible in the transformation of asset management, long a high margin and high service business, which came under immense pressure due to the innovation in the space. The launch of new competitors utilizing passive investment techniques, index funds and later ETFs put tremendous pressure on margins. Asset management and alternative investment boundaries would soon become blurred, with alternative fund managers expanding into the space left by traditional asset managers.

Finally, the emergence of new products related to the tranching and redistribution of risk attracted the interest of players in the insurance industry who viewed this space as a core competency. Examples for this are the now infamous AIG Financial Products division, but also the monoline insurers who branched out from their origins in municipal bonds to cover many forms of asset-backed structured products.

The macroeconomic forces acting upon the financial sector not only transformed the traditional space, but led to the creation of the modern day alternative investment ecosystem. The result was the rise of a previously obscure investment space led by family offices into the U.S.$6t alternative investment industry that exists today. The drivers of this growth can be traced to changes in investment laws as well as regulation aimed at the traditional financing space. In particular, the clarification of the prudent man rule, a part of the 1974 Employee Retirement Income Security Act, made it possible for pension funds to invest in privately held securities. This allowed alternative fund managers, such as private equity firms to re-orient their businesses from family offices and high-net-worth individuals to institutional investors – a trend that still continues.

The attractiveness of equity was subsequently enhanced by the 1981 Economic Recovery Tax Act, which reduced capital gains tax rates. In addition to making returns for institutional investors more attractive, this allowed alternative fund managers to improve their own economics by structuring investment vehicles to be taxed as capital gains and not as income. This gave the industry an edge in attracting high quality and innovative

Figure 8: Total debt of major economies
Source: Roxburgh et al. (2012)
Includes all loans and fixed-income securities of households, corporations, financial institutions and government
talent from other parts of the economy. Perhaps, an even more significant driver of change carried interest structures, conceived to address the principal-agent issue for long-term investment from passive (limited) principals to active (general) partners was used to boost returns under beneficial tax treatment.

The confluence of vastly improved technology, liquid capital markets and retreat of traditional active fund management lowered barriers to entry for new, technology-driven asset managers. Together with an appetite for yield by institutional investors and family offices, this enabled the rapid growth of the hedge fund industry to its current size of some U.S.$2.5t. For a long time, universal and investment banks with their “proprietary trading” units served as a training ground for hedge fund managers that would eventually start their own operations.

A (partial and incomplete) list of changing rules and roles in the financial services ecosystem

To illustrate the way in which the “rules and roles” in the financial services ecosystem changed, changing the architecture of the sector, it is worth identifying some of the actors which changed – often beyond recognition in the recent period. We focus on those that have become systemically relevant or that are new classes of actors.

Money market funds were created to overcome regulatory driven barriers that artificially constrained the deployment of short-term capital by the private sector. They have since become a critical node in the financial system, as many other players such as investment banks and hedge funds are heavily reliant on the short-term capital they provide.

Universal banks saw their structure and business model changed, with multiple sources of funding now being pooled, greatly complicating the historically linear picture of banks directly connecting savers and investors. The new model, with strong distribution networks already in place, supported the creation of a vast range of new financial products that served to further complicate the links between sources and destinations of capital. Increased derivatives volumes and use caused a marked increase in universal banks’ leverage and scale. The economies of scale allowed the largest universal banks to dramatically increase their share of the sector balance sheet, as can be seen in Figure 9; it also allowed investment banks to increase their scale and scope of activities and instruments. The highly leveraged nature of the newfound scale, the dearth of regulation governing much of the growth and increasing levels of interconnectedness within the financial system led to the point where “too big and connected to fail” became an issue. More importantly, the source of income changed drastically, as the analysis of bank holding companies suggests [Cetorelli et al (2012)]: revenue increasingly shifted from traditional (interest) income to fee-based and non-traditional income, and the number of activities, scale and complexity of these banks grew to unprecedented levels.

Rating agencies, arguably the most nodal and least effective part of the system, have found it remarkably difficult to be accurate, let alone far-sighted in their foray in the world of structured finance, which vastly expanded their volumes. Their serious shortcomings, and their inability to foresee the major financial catastrophes in the sector have been noted – a number of the financial services firms that were later forced to merge with other institutions or even went bankrupt were solid investment grade days before their bankruptcies. Their assessment of risk in tranches of CDOs was even worse, yet they have become systemically more important as regulation (especially through the Basel and Solvency frameworks) has turned them into de facto regulators. Governments recognized them as the official
benchmark of risk on record and regulation required many large institutional investors to utilize the benchmarks when selecting assets, restricting their ability to invest in ratings below a threshold. Well-intentioned, codifying the role of rating agencies into the financial system as these ideas might have been, they indelibly altered incentive mechanisms for asset managers, rating agencies, and wholesale banks. And given that the payment system was changed by the U.S. Congress in the early 1970s, making issuers and not investors pay for their services (while restricting competition to the small set of Nationally Recognized CRA), the rating agencies themselves had good reason to give in to the pressures to inflate ratings. In practice, agencies had little incentive to increase the quality of their ratings, but a large incentive to increase their volume; predictably, a huge increase in their business increased, and the revenue per employee shot up by multiples of three and above [Jacobides (2014)]. Furthermore, banks found it cost-efficient to reduce some of their own risk-assessment in as much as ratings could be substituted, with the net result being that far fewer people, with far less direct knowledge, were involved in real credit and risk analysis. The fact that rating agencies were aiming for increasing volumes of risk assessment in ever more opaque and complicated structure while cutting costs should have been a red flag. Yet, the fact that ratings agencies were all for-profit, profit maximizing public firms, or occasionally the cash-cows of corporate parents eager for results, did not seem to worry the regulators, at least not to the point of identifying an alternative. And despite recent legal action, their role in the center of modern finance has not been challenged.

Sovereign wealth funds (SWFs) have emerged as an important new source of alternative financing. The concentration of vast amounts of wealth by a finite set of actors, their foreign national identity, and the source of wealth (primarily oil and trade) have all helped to create the image of omnipotent foreign actors with the power to overwhelm domestic markets. The reality, in the context of the global financial system, is quite different. The assets managed by SWFs, while significant in absolute terms, are actually quite small relative to other actors such as pension funds, insurance companies, or private wealth managed by traditional asset managers. In addition to this point, Figure 10 also notes that SWFs play no intermediary role, further reducing their systemic relevance.

Alternative investors, are comprised in particular of private equity and hedge funds. Similar to sovereign wealth funds, they have outsized influence at the individual actor level, but as a whole the industry has not proven systemically relevant in the recent crisis. However, their rapid growth has made them an important actor and source of financing. Similarly, the industry simultaneously supports and relies heavily on the new financial ecosystem. Hedge funds often utilize money markets and wholesale banks for a material amount of their
short-term funding, serve as counterparties on large amounts of derivatives, and provide tremendous amounts of liquidity throughout the system. In turn, private equity and infrastructure funds make ample use of structured products in support of their capital structures.

Each major asset class seeks to generate value in a distinct manner, with possible positive effects on societal economic productivity. Early-stage private equity (venture capital) seeks to identify and support companies that develop and deploy innovative technologies and academic research finds that its impact on the economy has been far reaching [Lerner (2002), and Kaplan and Lerner (2009)]. Private equity seeks to enhance the productivity of companies by optimizing the operations, governance and capital structure of existing companies (Mozes and Fiore (2012)). A subset of private equity funds focuses on infrastructure and other physical assets. While their intrinsic value creation approach has changed little over the last decades, various adjustments to their business model are worth noting. As the industry matured, this segment became more institutionalized: while its roots were with high-net-worth individuals and family offices, today’s funds orient themselves toward institutional sources of capital such as pension funds and SWFs. Furthermore, the emergence of high-yield (junk) bond in the late 1980s dramatically increased the potential for private equity players to apply leverage to their deals and aim for ever bigger buyouts – a phenomenon that has seen several cyclical peaks since then. The ability to pursue large targets has also been enhanced by opportunistically collaborative behavior among buyout funds, which will regularly form consortia to increase their firepower. This strategic proclivity is not confined to acquisitions – an increasing number of private equity exits are made to another fund (so-called “secondary buyouts”), rather than into public markets. Finally, the increasing availability of structured products and financial innovation has allowed private equity players to create value through financial engineering – a trend that has somewhat abated since the 2008 crisis, but is nevertheless still relevant.

Hedge funds, which invest in existing assets, have been shown to increase market liquidity, encourage financial innovation, improve the governance of companies, and reduce the transaction costs associated with distressed debt (Blundell-Wignall (2007), Muhtaseb (2012), Fichtner (2011)). Starting from small origins, the sector has become a major force in capital markets – the 1998 LTCM episode being a case in point. This has been enabled by the changes in academic finance theory and technology discussed earlier, which have lowered both barriers to entry and barriers to scale. As a consequence, individual hedge funds can now account for a major share of daily trading volume on stock exchanges. Their size has enabled them to become market makers in addition to traders, and some funds run “dark pools” to that effect [Farrell
In addition, hedge funds have broadened their business model from origins in “stock picking” to cover the whole range of financial assets – major funds are now active in commodities, credit, foreign exchange, and all other asset classes. Hedge funds have also been at the forefront of the recent expansion of algorithmic and high-frequency trading, which both capitalizes on and increases the connectivity of global markets.

Central banks have long been key players in the financial system, but only in the recent crisis have they become “actively systemic” again. They have moved beyond their traditional, narrow mandate (manage economic stability, with inflation and possibly employment in mind; manage money supply and exchange rate; and acting as a lender of last resort during times of crisis). They have shifted from their role of a “lender of last resort” (promise to pay) to “liquidity provider and market backstop of last resort” (promise to buy) that ultimately elevated central banks to their preeminent position they occupy today. Their transformation can be seen through the changes in the balance sheet – especially of the Federal Board of Reserve ([Mehrling et al. (2013)]).

The 2008 financial crisis as an illustration of systemic interactions

Having seen the changes in the finances services ecosystem, we can turn to the examination of how these interact, using the recent crisis as a canvas.

The seeds of the most recent crisis can be traced to a fragmentation of the regulatory climate in the U.S. surrounding mortgages and a secular decline in the mortgage lending standards that regulators allowed. Nobody in the regulatory context viewed their role as a guarantor of the upkeep of sensible lending standards at the retail end. Even discussions of control over predatory lending (which had very real social implications) were met with hostility, leading to fairly ineffective regulations. But, fundamentally, the elementary questions of “who gets loans, and should they? And who bears the risk if they default?” were not addressed head-on. The fragmented
nature of lending regulations in the U.S. made it possible for many of the leading mortgage originators to avoid Federal oversight. At the same time, the Office of Thrift Supervision was subject to regulatory capture by the firms that it was tasked with overseeing. And while it was poor loan originations in the U.S. that created the crisis, investors in such loans were globally diverse. Major investors, such as German banks, were not only poorly supervised, but also incentivized to maximize their returns subject to the risk assessment of the ratings agencies. So, weak supervision of depository institutions in other geographies would eventually fuel the demand of loans of declining standards. Sadly, nobody was looking at the system as a whole, or at the entities they supervised with a systemic mindset.

The transformation of the financial system had altered the structure of the system, and with it the distribution mechanism for new products. The new scale and connectedness of universal banks and the links between them, investment banks, and the capital markets provided a global network that could distribute risk around the globe to a degree and at a speed that had previously never been seen. Moreover, inadequate regulatory models allowed them to reach record leverage levels. Large banks in the U.S. saw their leverage reach 25x–30x by 2008, while their peers in Europe mirrored the change, with leverage ratios even higher at 30x–50x in the same period (partly due to different accounting treatment of derivative positions rather than a fundamental business model difference). Figure 11 illustrates the point, while also showing the effect of an assumed government guarantee for certain intermediaries (government sponsored entities or GSEs).

The arrival of structured products and the (believed) analytical capability to accurately price them led to a dramatic expansion of debt. It really was too good to be true – transforming high risk securities into investment grade securities, with the blessing of ratings agencies, attracted a large pool of investors into low-grade debt. Simultaneously, the belief that those assets were truly lower risk allowed more debt to be issued from a lower capital base. Predictably, the result was a rise in debt issuance for a wide variety of non-investment grade debt, with issuance of ABS, MBS, CLO/CDOs, high-yield bonds and leveraged loans all increasing dramatically.

Asset prices strayed far from their historical averages, with real estate prices in the U.S. and Europe leading the way, soaring 50%–100% from 2002 to 2008.

The stage was then set for the crisis. In the summer of 2007, two hedge funds specializing in mortgage-related securities and

---

**Figure 14: Prices of U.S. mortgage-related securities, 2006–08**
Sources: JPMorgan Chase and Lehman Brothers. (ABX = an index of credit default swaps on mortgage-related asset-backed securities; MBS = mortgage-backed securities)

<table>
<thead>
<tr>
<th>2006</th>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prices of U.S. mortgage-related securities (USD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABX BBB</td>
<td>ABX AAA</td>
<td>Alt-A</td>
</tr>
</tbody>
</table>

**Figure 15: Prices and spreads for select debt instruments**
Source: Bloomberg

<table>
<thead>
<tr>
<th>2007</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sovereign and others: seven countries, 10-year U.S. Treasury</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>U.K.</td>
</tr>
</tbody>
</table>

Prices of U.S. mortgage-related securities (USD)
run by Bear Stearns suffered tremendous losses when markets began to formally recognize the mispricing of real estate assets. Financial institutions and asset managers across the world marked down related securities, leading to reduced collateral levels, capital calls and ultimately a liquidity crisis. What began with mortgages would lead to a severe shortage of credit, bring new debt issuance to a new halt, and expose the risks associated with off-balance sheet vehicles and SIVs (Bear Stearns had to bail out the hedge funds in question.)

Unexpected losses in subprime mortgages led investors to fundamentally reassess their models and assumptions. The value of securitized products fell precipitously (see Figure 14) and spreads on different grades of debt widened dramatically, before returning to historical levels (see Figure 15).

The efficiency of the highly interconnected system began to work in reverse, rapidly transmitting losses and destabilizing the system. The degree of leverage that had hitherto supported return on equity now served to undermine entire institutions, as thin equity cushions were rapidly worn away. The crisis exposed the heavy reliance on short-term funding by wholesale institutions and left them particularly vulnerable to capital flight, as shown in Figure 16.

Bear Stearns provides a perfect example, losing virtually all its liquidity buffer in the course of a single week and being subsequently acquired by JPMorgan as a result (see Figure 17).

AIG turned out to be a hitherto unrecognized “super node” of the financial system. The use of derivatives contracts by an obscure London based unit of the New York-based insurance giant (inadequately regulated by the Office of Thrift Supervision, a regulator without extensive derivatives expertise) tied it to a wide range of other financial institutions across the globe. The underlying assets tied to the insurance company to the mortgage industry. Losses in the mortgage industry led to losses on the derivatives contracts, which led to margin calls by AIG’s counterparties. The sudden insolvency of AIG as a result of the losses carried the immense risk of dozens of key financial institutions losing their counterparty. The collapse of such a systemic node would have proven catastrophic to the global financial system, forcing the U.S. government to bail out AIG.

The failure of Lehman Brothers, though not as interconnected as AIG, sparked a wholesale reassessment of counterparty risk, further exposing how complex and interconnected the financial system had become. Virtually overnight, many markets across asset classes ceased to operate entirely. The sudden and unexpected loss of liquidity exposed modern portfolio management theory, the bedrock for the allocation of
The Journal of Financial Perspectives

Rethinking the future of financial services: a structural and evolutionary perspective on regulation

much of the world's capital, with its reliance on diversification among uncorrelated assets. Virtually, all assets became highly correlated, exacerbating dynamic feedback loops and the severity of the crisis.

Congressional testimony in September 2008 noted that White House and Federal Reserve officials feared the outright collapse of the economic system of the U.S. due to a run on money market funds [Jones (2009)]. Such a scenario would have been possible if losses in mortgage backed securities would be transmitted into what investors had believed was a near zero risk substitute for traditional interest generating time deposits. The FDIC was forced to launch the Temporary Guarantee Program to avert the crisis.

By the fall of 2008, governments around the world began to realize the potential for another Great Depression and quickly took unprecedented actions to head off such an event. Their response is well documented and we will not elaborate further here, but refer the reader to US Department of the Treasury (2013), Lannoo (2011) and Laeven and Valencia (2012) for the major financial markets. The response, led by the G20, consisted of nationalization of major banks, and support via government stakes and access to liquidity for those that stayed independent; bailout of major non-bank financial players; significant Keynesian stimulus to the economy; extension of guarantees and other confidence-boosting measures. Not surprisingly, the response led to a major increase of sovereign debt levels and central bank balance sheets in the OECD economies.

Where we are: regulatory initiatives in overdrive, system in flux

With contraction and indebtedness dominating the public agenda, politicians determined that “the system had to be fixed,” and demonstrating a tough approach to the industry was inevitable. There has been a mad flurry of regulations, mostly in the U.S., the E.U. and the U.K., which have surely increased the administrative requirements on businesses and focused on piecemeal aspects of the sector, even if major issues remain substantively unaddressed. The question of whether we have focused on volume and vindictiveness of regulation rather than solving the problem remains. Regulatory reforms have been mostly aimed at identifying systemic actors, reducing their risk profile and increasing transparency; and their side-effects have not been worked out yet.

The majority of financial reforms focused on the banking system. Collectively, the Dodd-Frank Act in the U.S. and the Basel III, Capital Requirements Directive IV (CRD IV), and Solvency II acts in the E.U. serve as the primary vehicles for change. They seek to increase leverage, capital, collateral and liquidity requirements, and limit the ability of banks to engage in proprietary investing or trading. The overall objective is to reduce the risk profile of financial institutions and to make them better able to withstand shocks in the future.

Regulators are also seeking to increase the level of transparency in the financial system. The E.U. and U.K. driven Packaged Retail Investment Products (PRIPS), Undertakings for Collective Investment in Transferable Securities V (UCITS V), and Alternative Investment Fund Managers Directive (AIFMD) laws and the Foreign Account Tax Compliance Act (FATCA) law in the U.S. are key initiatives. The laws require investment firms to register with authorities in their jurisdiction, provide detailed financial and operational information on an on-going basis to regulatory authorities, share detailed tax information on clients, and make available materially more information to the public than was the case in the past.

Regulators are keen to better define the nodes of the financial system, particularly in derivatives. Dodd-Frank, UCITS V, AIFMD, European Market Infrastructure Regulation (EMIR), and the Markets in Financial Instruments Directive II (MiFID II), require financial intermediaries to clear most derivative contracts on centralized exchanges and investment firms to utilize independent depository firms. The goal is to reduce counterparty risk at the systemic level and to reduce fraud at the investor level.

Compensation and incentives are one of the areas where cross-border synchronization has broken down completely, as U.S. regulators have elected to refrain from any related rules. In contrast, the E.U. and U.K. have introduced new guidelines on both quantity and structure of remuneration.

While all these efforts are well intended, their sheer number risks a proliferation of, at times, incompatible rules with high
compliance costs. Furthermore, many of those rules will carry secondary effects that are not yet clear; Figure 18 provides our first cut at what they may be.

Ironically, most of these efforts have been partial and fairly segregated. They have focused on particular institutions, slivers of the financial services sector, without considering the systemic issues that emerge. For instance, even the major jurisdictions summarized above (the U.S., the U.K. and the E.U., which together hold the vast majority of global banking assets) have advanced differently, and often competing regulations, leading to a host of cross-border regulation issues. Also, regulations aimed at financial resilience have led to serious credit contraction for some groups of customers — especially the SME and mid-sized corporations, while favoring sovereign debt, as we note further on. Most importantly of all, the role of the rating agencies, deeply embedded in the web of the financial services world yet still privately held, profit maximizing public listed companies (or subsidiaries of corporates) has not been addressed.

Post crisis, regulation has been piecemeal, focusing at one problem or institution or facet at a time. Ironically, the excessive fragmentation which (applied to mortgage banking) led to the collapse of the financial system is making a comeback in leading to a fragmented regulatory agenda, with little understanding of its systemic and unanticipated consequences. All the while, the sinful practices of easy lending that it was intended to cure are starting to re-emerge, be it in high LTV mortgages or covenant-light business loans, are becoming ever more present as a result of easy money from central banks.

A structural evolutionary approach to the financial system

Our analysis above provides an illustration of what exactly a systemic analysis can offer, and how the sector’s structure could, predictably, lead to this evolutionary disaster. The problem was not that there was some exogenous increase of greed or incompetence, but rather that the system allowed people to focus on their own limited part, which led to a collective disaster. As Rajan (2011) put it, “each of the actors – bankers, politicians, the poor, foreign investors, economists, and central bankers – did what they thought was right. Indeed, a
very real possibility is that key actors like politicians and bankers were guided unintentionally, by voting patterns and market approval respectively, into behavior that led inexorably toward the crisis.”

The problem, then, was that there was little understanding of the dynamics of the system as a whole. Even worse, as just discussed, more attention is being focused on each of the individual components of the system, potentially working to destabilize, rather than redress the risks and imbalances. And even today, we may already be building the next bubble, as lending practices are starting to converge to their risky pre-crisis structures: covenants are reportedly becoming ever lighter, and with excess liquidity in the system and few deals competing to absorb it, risk may be creeping back into the system.

It is remarkable that in most economic models, there still is “one central bank” with “one interest rate”, and a banking sector that is nothing like the complex institutional set of arrangements that we see today. Even more so, that there is little effort to try and comprehend the institutional complexity and the regulatory challenges it produces. Regulators are often unaware of the business structures they regulate, and there is no locus of exchange of opinion on systemic dynamics in the financial services sector, leaving any opportunities of jointly overcoming these issues to waste. The engagement of different players is, sadly, predictable: regulated entities try to fend off regulation, whereas regulators try to assert their authority, suspicious of their regulated entities. The possibility of a regular forum for exchange of views of what the issues are, which would include those providing capital, those needing it, and sets of the intermediaries and their (often manifold) regulators is absent, be it on a national or international level. The FSB has a nominal set of 78 “systemic” priorities at the time of this papers’ writing, predictably very few of which progress.

What we advocate is an institutionally grounded, business-model focused approach, which recognizes the different species in the financial services ecosystem, considers their evolutionary dynamics, and intervenes to manage risk. In this sense, we feel that there is significant mileage in leveraging the recent work of Andrew Haldane and others, who laid forth a vision of the financial system as an ecosystem akin to that found in the natural world (Haldane (2009)), as well as the rich tradition of evolutionary economics (Nelson and Winter (1982)), which has spawned a large research stream which we may usefully draw on. There is little gainsay on the value of mapping and understanding the topology of the financial network; system-wide properties such as connectivity, diversity, and complexity are clearly important. But we think that we need to complement this approach by a better understanding of the business logic of the different participants. That is, we would argue that we need to consider the specific distribution of rules and roles in the sector, of the nature of agents within it, and how this shapes the propensities of the actors involved (Jacobides and Winter (2010)).

Figure 19: Beneficial effect of small, controlled burns in forestry management
Source: Piñol et al. (2005)

We are aware that systemic analyses have a bad name in their propensity to delve into discussions of complex systems that provide little practical guidance. We do, however, think that this need not be the case, neither in principle (as we will discuss in this section) nor in practice (as we will argue in the next).

A systemic approach could manifest itself more than as a set of institutional mechanisms. It would allow us to be more critical of some of today’s attitudes of policy makers, including a strong desire to eliminate “negative” financial actors (e.g., short sellers); narratives that emphasize only positive outcomes — “banks will never be bailed out again”; or an emphasis on “safe” and “simple” business models (when viewed in isolation), as if the rest of the system did not exist.
We note that little attention is paid to the financial system’s equivalent of biodiversity — diverse business models that neither rely on the same input nor pursue identical strategies. An example of reliance on the same input was the systemically dangerous funding model of the majority of the banking sector on short-term wholesale markets. An example of identical strategies are “crowded trades” leading to imbalances and asset bubbles (e.g., the yen carry trade for most of the 2000s).

Top-down regulatory approaches tend to bring about such dangerously homogenous strategies as they try to force business models into supervisory buckets that can be easily understood and monitored. The natural response of actors will be to optimize the business model against the static regulatory parameters, which over time will lead to homogeneity.

Cumulatively, this stance might well be the equivalent of the “zero fires” approach in forestry that is now discredited (see Figure 19) — financial crises that happen less frequently, but with much greater intensity [see World Economic Forum (2010) for further parallels in risk management between the financial system and other domains]. The recent initiative to introduce stress-testing exercises for banks and systemic financial actors is welcome in this context — however, it remains to be seen whether those largely model-based exercises can fulfill the function of a “controlled burn” or will simply end up being another game-able regulatory parameter.

As we consider how we can best draw on, and leverage the lessons from ecosystems [e.g., Haldane (2011)], we need to be mindful of the role of structure and recall that human actors are unique in their ability to not only optimize their personal outcomes within a given system, but they also have the power to shape the system itself (of which lobbying is only one example). The emergence of the securitized mortgage banking sector was the result of endogenous pressures to change from actors who stood to benefit from this [Jacobides (2005)] — a theme also pervasive in Tett’s account of the derivative market growth [Tett (2009)]. If actors are continually seeking to alter the ecosystem to their liking, then regulators too must continually adapt to the changing state of the system. In such a world, static rules will naturally depreciate with time; the idea of “building a safe system from now onwards” is nothing short of self-delusion.

We argue that we should move beyond the application of the analogy of an ecosystem, which has been, by and large, limited to a topological discussion of “banks” as “nodes” and of the role of “interconnectedness”. We consider the crucial role of the changing division of labor — the changing industry architecture of the financial services world, but also the changing architecture of the rules, responsibilities and remits of the regulators, which might be sowing the seeds of the next, likely crisis, by its inadvertent systemic side-effects. We claim that a better understanding of the structure of the financial services sector,
Rethinking the future of financial services: a structural and evolutionary perspective on regulation

not only in terms of the regulated part, but also on the “shadow banking” side, and of the role of other providers of capital, is crucial. And we believe that a full mapping of the activities following [Pozsar et al. (2010)], is a critical part of the analytical and empirical foundations of the sector. In other words, we feel that an evolutionary approach, which concentrates on the incentives and propensities for action in the financial system, is a critical element in both understanding the sector (and what problems it might bring about) and regulating it.

In order to address these shortcomings, two steps must be taken. First, regulators must draw up historical and present day maps detailing the entire global financial system focusing on the evolution of institutional structure. An emphasis would be placed not only on looking at firms as nodes; but also understanding the business models, and the incentives within them. In this regard, the recent mapping by the U.S. Federal Reserve of the traditional and shadow banking systems [Pozsar et al. (2010)] is a solid step in the right direction.

The second step would entail analyzing this data in order to understand how the system has evolved over time, and what the underlying economic rationale for the evolution is. Critically, such analysis should not be based solely on economic theory, but should also seek to identify and track the runaway growth in business models as a key regulatory metric. This would comprise credit levels naturally, but would also include crowded trades or systemically relevant product innovation – and, for example, would have identified AIG as a systemic node well in advance of the crisis, giving regulators the opportunity to step in before the actual crisis had hit.

Looking for business models exhibiting runaway growth can also help with another frequent issue in the financial system – incentives and compensation. While genuine innovation should reward the innovators, rent-seeking should be discouraged in a system that underpins the functioning of society at large. Monitoring business models means innovation turned rent-seeking can be spotted before it becomes a problem. A variety of interventions can help break the self-reinforcing cycle of the best talent being attracted to such potentially extractive areas, of which regulatory intensity could be one (see Philippon and Reshef (2012) for a discussion of the issue in the financial system at large, and Figure 21 as an example of their findings).

Rethinking the current regulatory landscape in a dynamic system

With this background set, we move to an intentionally provocative section that revisits both the background and state of affairs in terms of the regulation (including our view of its shortcomings, seen from a dynamic perspective). We conclude with speculative recommendations for regulatory action, hoping to stir debate on the topic.
Historical background

In order to craft proposals aimed at strengthening the regulatory system, the structure and limitations of the existing system must first be understood. Prior to the financial crisis, the global financial regulatory system had evolved into a collection of agencies that largely specialized in a particular geography and vertical line of business (see Figure 22). Such a system would indeed make sense in a traditional financial system with limited spillovers between verticals or geographies. However, such a structure would not work well if the system was both dynamic and interconnected (as we have shown the modern financial system to be). Alas, that is precisely the world regulators were tasked with overseeing prior to the financial crisis. The increasing size and interconnectedness of the nodes that made up the financial system limited the ability of actors and regulators alike to track or understand the risks that were building up [Haldane (2009)]. Concomitantly, the increasing complexity of the system and the products that it used also proved too much for the early warning system in the private sector, clouded by rating agencies’ benign risk assessments. The focus on specialization proved highly ineffective, as the narrow purview of regulators left them unable to see how the system as a whole was accumulating unsustainable levels of risk.

Furthermore, the political context within which regulation has happened is fraught with adverse incentives. Politicians, wanting to show themselves spendthrift and decisive in times of crises, have often cut back on the budgets and pay levels of regulators much as the demand for supervision and the skill required to supervise has increased. Given the lack of sympathy banks and bankers elicit, there are some real pressures, only partly counterbalanced by the ability of major financial institutions as campaign contributors or providers of credit. The result of this difficult balance, and of the relative paucity of means of regulators, amplified by the bureaucratic reflexes of governmental organizations has been to propose new regulations, whose aims are not only to deliver a safer financial system, but also to show to the voting public that politicians are willing to “take on” the banks, often creating a series of administrative burdens with very real costs (often passed on to final customers, as price increases or credit availability decreases), yet debatable effectiveness in stemming a potential crisis.

This trend has had another, less understood side effect: some estimates on the advisory work needed by banks to adjust to new regulators exceeds U.S.$4b, and often the advisors who help banks comply consist of former regulators, who might have been paid scantily in their previous positions but are able to benefit from the largesse of their new employers. The major accounting and advice firms as well as many consulting firms have ramped up their hiring of former gamekeepers, who, seeing their salaries decrease, have responded to the market needs. This has been leveraged by regulatory authorities, most notably the SEC, which now regularly asks firms to “seek advice” (from a private sector consultant) when it is “concerned” with some practices or compliance issues.¹

Against this backdrop, the global regulatory community has been actively engaged post-crisis in seeking to establish new guidelines to prevent a similar crisis from occurring. In doing so, they must work within the framework established by their governing political system, which is subject to pressure and influence by private actors and firms, as well as with the objectives of the real drivers of regulations – politicians, who face a host of conflicting, and mostly shorter-term objectives and often have a limited understanding or patience with this complicated and unpopular sector. Within that context, regulators, whether by choice, inertia or dictum from their

¹ The U.K. appears to follow the example in the U.S., in terms of private firms use for public mandates, having just established the “Section 166 Reviews”. These reviews, to be required by the regulator, are to be done by a small number of pre-selected private companies, whose role is to satisfy the regulators they should not intervene. While this preserves public funds, it does raise questions on the governance of oversight.
political superiors, continue to focus their efforts on controlling institutions within their defined vertical and “playing God.” In doing so, they (and their political superiors) risk both not resolving the institutional bases of the crisis and not fulfilling their role of ensuring system wide stability.

We identify eight fundamental fault lines in the present regulatory structure that we believe should be addressed in order to strengthen the ability of regulators to ensure a stable and sustainable financial system.

- **Rules assume that regulators possess knowledge they do not and cannot have** — the regulatory approach is highly formal and process driven. Its reliance on static metrics and one-way reporting by firms is inefficient at best and reduces the ability of regulators to react timely to changes in a dynamic system. The benefits to the system are questionable, but the compliance costs are clear. Financial institutions have committed to hiring thousands of new compliance employees. The trend is reflected at the industry level, with hedge funds alone having spent more than U.S.$3b on compliance since the crisis [KPMG (2013)].

- **Rules assume that regulators have powers they do not and cannot have** — the nimble nature of the private sector, coupled with the talent that it can attract and remunerate, leaves regulators at a structural disadvantage when supervising such actors. Regulators should jettison approaches that rely on an illusory paradigm of omnipotence — the alternative of trying to hamstring the private sector is unlikely to work in practice.

- **Current metrics are inherently insufficient to prevent future crises** — present metrics focus on analyzing static balance sheets, rather than the dynamic and market-based view that would be needed to measure and mitigate stresses in real time. In today’s complex financial system, a static “clean” snapshot may be the last view regulators have before another crisis comes from an unexpected corner and rapidly engulfs the entire system.

- **Regulatory framework relies on a classification methodology that is no longer relevant** — basing regulations on fixed verticals associated with well-defined and relatively exclusive roles, and populated by mostly pure play actors makes perfect sense in the traditional financial system. For better or for worse, the modern day financial system is no longer so well organized. Examples of firms in one vertical engaging in activities traditionally associated with another vertical include: (1) private equity firms, hedge funds, and asset managers issuing credit and loans; (2) hedge funds providing reinsurance; (3) insurance firms investing in and serving as counterparties to derivatives contracts; and (4) universal banks simultaneously engaging in retail, wholesale and asset management activities, and interfacing with the shadow banking market.

- **Regulators use tools that are no longer appropriate to manage the system** — regulators seek to define input constraints (capital) and operating constraints (licenses) based on classifications of industry verticals. This no longer matches a dynamic ecosystem where business models change according to market opportunities. The fixed capital and liquidity requirements found in Basel III and Solvency II are good examples — like their predecessors, they are likely to invite “crowded trades” by players trying to optimize the rules for themselves. This recognition, which may have underpinned the recent re-orientation of regulators to control leverage, might substitute one blunt measure (which can be perverted by market participants playing with the rules) with another.

- **Regulatory fragmentation on a geographic and segment level continues** — the global and highly interconnected nature of the modern financial system requires that regulators coordinate their actions and rules to a high degree. Unfortunately, national political concerns result in unique regulations and guidelines pertaining to such critical issues as cross-border resolution requirements, remuneration guidelines, and capital or liquidity requirements. At the very least, a balkanization of the financial system would decrease efficiency of capital allocation. At the worst, systemically threatening imbalances could result if actors actively sought to arbitrage jurisdictions at scale.

- **Regulation maintains an unhealthy link between banks and sovereigns** — as a remnant of an older regulatory philosophy, regulation primes sovereign debt, especially in terms of bank balance sheets, by zero-risk weighting. This means that, both for capital and also liquidity reasons, banks tend to rationally hold increasing amounts of sovereign debt, to the detriment of other forms of loan. This creates an unhealthy link, whereby banks are exposed to sovereigns’ risks (as evidenced in the Cypriot banking crisis); and also where sovereigns build dependencies on banks, which absorb their own sovereign debt when it is issued, turning us back to the Medici era of financiers supporting the state. Unfortunately, with many countries (especially in Europe) having both sovereign and bank debt issues, the resolve to sever this unhealthy link appears limited.
Rating agencies are still as nodal and as problematic as before — while rating agencies have been shown not to be effective during the last crisis, and it has become obvious that their payment and governance structure would make it unlikely that they would change themselves effectively, little has happened to remove their quasi-regulatory license. We still have a system that depends more than regulators would like to acknowledge on the links between regulated and non-regulated intermediaries, with the capital markets being essential to its function; yet the gatekeepers and regulatorily mandated assessors of risk still have not been properly restructured. Even their governance structure (publicly listed companies or subsidiaries, as opposed, e.g., to partnerships, or structures with some liability in the model of auditors who also make representations) has not advanced.

The fault lines in action: inadvertent consequences
The complexity of the modern financial system and the breadth of reforms being enacted simultaneously has notably increased the risk that the law of unintended consequences will strike both the traditional and alternative aspects of the financial sector (as Figure 18 showed). The ability to target reforms at a particular actor group and isolate the effects may hold true in the old static model of finance, but it no longer holds true today. In the modern system, regulators must consider how a rule will impact not only the target group or function, but also all of the other parts of the system that are connected to it.

The current slate of financial reforms, while quite beneficial in many respects, runs the risk of unnecessarily slowing the pace of the economic recovery due to their inadvertent effect on areas beyond the target of the reforms. We give four examples of such unintended consequences in the traditional financial sector.

The current U.S./E.U. capital and liquidity guidelines (as laid out in Basel III, Solvency II, Dodd-Frank, and updated IFRS accounting standards) incentivize firms to hold highly liquid securities. In practice, this means prioritizing sovereign and mortgage debt over riskier loans to SMEs and infrastructure projects. The benefit to society of supporting job creating SMEs and investing in much needed infrastructure is clear, but the current regime serves to undermine both intents.

Another key area of concern is the effect of regulatory differences in cross-border resolution regimes. A certain tendency of national regulators to ring-fence local operations of global banks might create “capital traps” that lead banks to either shrink their footprint or increase prices to compensate for the inefficient allocation of their resource. While it is understood that cross-border resolution regimes are difficult, and therefore the regulatory reaction to atomize the financial system is understandable, a delicate balance needs to be struck here.

The introduction of government mandated remuneration structures in the financial sector in the E.U. and the U.K. is another delicate point. The goal of better aligning incentives for individuals with broader societal goals is admirable. Unfortunately, the policies might drive talent into the unregulated sector and away from the parts of the system where they could provide innovation needed by society (e.g., in infrastructure finance).

Proposals by the European Commission to introduce a financial transaction tax on most types of securities trades would hit the repo market, on which the majority of banks depend for their overnight funding, especially hard. The International Capital Markets Association and the European Repo Council of ICMA recently warned that the number of repo transactions could fall by 50%-66% [Stevenson (2013)] if the original proposal by the E.C. was approved and implemented. Both the U.S. and the U.K. have expressed concerns, fearing severe ramifications for their domestic and the global economic system [Atkins (2013)].

The impact of post-crisis regulatory reforms is also being felt by the non-traditional financial sector, both indirectly through the large number of connections to traditional players, and directly as the industry now increasingly falls under the regulatory umbrella.

The reforms will force the industry to become far more transparent, which is in principle a good thing. However, second-order effects could also lead to a reduction in the level of innovation in the sector, as well as the ability to make the patient investments society requires to generate wealth in the long term.

As banks shrink their balance sheets in order to comply with new capital regulation, the level of leverage they can extend to alternative fund managers, such as private equity is curtailed. This in turn reduces the ability of those players to make acquisitions and (if their value proposition is accepted) optimize the operations of companies.
Similarly, increased transparency requirements for many alternative asset managers might lead to a similar short-term focus already observed in other parts of the financial system and the corporate world, to the detriment of longer-term investment choices – probably not what rule makers would like.

Some might argue that fewer private equity deals is exactly the right answer to the crisis. Before jumping to that conclusion, we submit that society (e.g., through pension funds that invest through hedge funds and private equity) does benefit from the innovation and intelligence of the non-traditional financial sector – and many activities of the sector respond to direct needs of the real economy.

**Implications and recommendations for regulators**

There is no question that regulation must exist to ensure the stability and proper functioning of the financial system with as few “unintended consequences” as possible. Since the financial crisis (and before), many well-intentioned regulatory initiatives have sought to improve the stability of the system. While we do not want to belittle those at all, we nevertheless believe they have been to a large degree reflecting a mix of Newtonian and Olympian philosophies, namely a reliance on the illusion of an omniscient and omnipotent regulatory body that can govern based on a set of fixed rules. Such an approach is in our view bound to invariably fail again as it cannot cope with a dynamic system comprised of intelligent, self-interested actors.

Transitioning to a dynamic system can be gradual, as new regulations (or the removal of old ones) are best made incrementally. In this spirit, we offer nine recommendations on how regulators could improve their ability to ensure a stable system.

- Increase “biodiversity” by lowering barriers to entry for innovators – doing so will allow new entrants to emerge and pursue the potential excesses in the market, thereby naturally reducing extractive or rent-seeking behavior. Specific examples where this approach could be applied currently include: (1)...

- Create a map of the shifting architecture of the financial services intermediaries – while it might be challenging to undertake, it is essential to understand the institutional layout of the sector, which would then allow us to assess both the incentives and potential pathologies of current institutions given the actual business models in place, and not the assumed interest of “a bank,” and also understand which actors want to change the architecture of the system, as well as the potential risks for doing so. This should include both the regulated and unregulated parts of the system – since it is in the interactions between the two that the potential risks (or solutions) may lie.

- Upgrade regulators’ talent and understanding of the private system – regulatory staff currently consists of highly skilled individuals trained in economic theory and highly trained auditors knowledgeable in applying complex sets of rules. Regulators would benefit from staff with a deep understanding of the business models employed by the firms they are overseeing, how they generate profits and evolve. Unfortunately, such individuals are currently scarce inside regulators. Having them would make for a more constructive dialogue with regulated entities, both sides having a closer match in vocabulary.

- Monitor runaway business models rather than static financial parameters – regulators would benefit from teams monitoring the industry for business models and products that exhibit rapid growth out of line with the underlying economy. This will allow them to spot imbalances before they become systemic. To achieve this, they need to have more staff with a business strategy background, as well as work closely with the business development and product approvals units of the institutions they supervise. (This is analogous to the WHO working with local virologists as well as monitoring public sources for keywords that could signify a disease outbreak.) Once a runaway business model is identified, it can be dampened with a combination of: (a) increasing input constraints (e.g., higher capital requirements); (b) attracting alternative competing business models into the space; and (c) constraining growth/volume outright. This point is a pre-requisite for the following two additions to work properly.

- Increase “biodiversity” by allowing negative market signals – given the complexity of the financial ecosystem, it seems highly inappropriate to restrict available market signals or business models only to those that convey a “positive” message. In this spirit, short sellers should be viewed as highly useful indicators of potential fault lines as well as “stressors” that can increase the resilience of the financial system. Similar considerations apply to other business models that are the equivalent of “predators” in natural ecosystems with highly beneficial effects in that context.

- Increase “biodiversity” by constraining growth/volume outright (b)
further enable crowd funding/P2P models to enter the startup lending space; (b) avoid the vertical oriented regulations and allow non-banks to actively engage in SME lending; and (c) relax licensing requirements for financial firms below a certain scale, to be reviewed once these firms grow.

- **Use arbitrageurs to stabilize the system** – rather than decreeing regulatory arbitrage away through presumed omnipotent fiat, regulators should create a (confidential) repository where all such arbitrage needs to be logged without reprisal as long as the arbitrage is legal. Regular review of this repository will allow regulators to fix systemic loopholes. As a side note, this is analogous to the computer security industry employing hackers to test their software.

- **Regularly test the relevance and applicability of regulations** – regulations should have a fixed shelf-life (say, 3–5 years), after which the continued validity and applicability is tested with real-world feedback. The tests should focus on usefulness to the economy and financial system as a whole, rather than on static definitions/interests of verticals or actor groups. Creating such a dynamic and flexible structure would allow regulators to better mirror the ecosystem and reduce the risk for system-wide failures.

- **Create a non-partisan platform for regulators and financial actors to discuss the financial system** – the financial system and society at large would benefit significantly by establishing a permanent forum through which all key stakeholders in the financial system could regularly communicate and engage with one another. The Financial Stability Board might prove an adequate platform for this. Such dialogues would enable all participants to deepen their understanding of the constantly shifting state of the financial ecosystem and thus be in a better position to maintain its stability and prevent a future crisis.

- **Engage key global regulators in events-based tests and war games** – regulatory tests could vastly improve if they utilized an events based approach with a “red team” actively trying to break the system, in addition to stressing economic model rules. Engaging in war games with fellow regulators (and industry players) would help supervisors better predict how different actors might behave in a crisis. It would also serve to identify previously unknown or unexpected issues, which will in turn allow regulators to design interventions. The BIS and Nordic regulators have reportedly engaged in such exercises, and such workshops of a more or less informal nature do take place. Yet we think that a scenario-based resilience check may be much more important, and needs to be much more mainstream than the ones currently in effect. This would be a large-scale change in terms of the way regulation is practiced, locally and globally.

**Conclusion**

As the complexity of the world around us changes, it is entirely understandable to hark for a simpler world [Haldane (2012)]. Yet, the quest for a simpler past is likely to be nothing short of a chimera. All the while, the regulated and unregulated (or differently regulated) parts of the financial system are coevolving in new, potentially unstable ways. Our paper suggests that we would be well advised to change our vantage point and analytical tools as we engage with the sector. A firmer base in the business models, the evolving industry architecture, and the dynamic properties, seems overdue. We need to shift from the study (and regulation) of individual “species” in the Financial Services ecosystem to a dynamic analysis of the messy reality firms inhabit. We hope that this paper, selective in its short space and provocative in its approach, will be a step in this direction and will engender further debate and discussion.

**References**


Bernstein, P. L., 1992, Capital ideas: the improbable origins of modern Wall Street, Maxwell Macmillan International


Rethinking the future of financial services: a structural and evolutionary perspective on regulation


Farrell, M., 2013, “The computers that run the stock market,” CNN Money, 8 July


Graeber, D., 2011, Debt – the first 5,000 years, Melville House Publishing

Haldane, A., 2009, “Rethinking the financial network,” presented at the Financial Student Association, 28 April in Amsterdam


Holters, J., 2012, Hedge funds and financial markets, Gabler Verlag


Jacobides, M. G., 2014, “Rethinking the financial crisis: structuring our historical understanding of a predictable evolutionary disaster,” forthcoming, Business History


Jones, S., 2009, “The formula that felled Wall St.,” Financial Times, 24 April


Lim, J., 2013, “The role of activist hedge funds in distressed firms,” working paper


Muhtaseb, M. R., 2012, “Hedge funds are the alternative mainstream,” Journal of Investing Summer, 57-60


Philpott, J., K. Beven, and D. Viegas, 2005, “Modelling the effect of fire-exclusion and prescribed fire on wildfire size in Mediterranean ecosystems,” Ecological Modelling 183, 397-409


Rooff, K., and C. M. Reinhardt, 2008, This time is different: eight centuries of financial folly, Princeton University Press


Stevenson, A., 2013, “Bank funding threat from EU tax,” Financial Times, 8 April

Taleb, N. N., 2012, Antifragile: things that gain from disorder, Random House

E.U. Financial Transaction Tax: a concern for all members of the C-suite

David Williams
Partner - Banking and Capital Markets Advisory, EY LLP

Mark Persoff
Partner - Banking and Capital Markets Tax, EY LLP

Abstract
In the last two years, French and Italian governments have introduced new taxes on financial transactions. A proposal for a broader E.U. Financial Transaction Tax exists, and its design could bring global financial institutions into its reach. The design of the tax will evolve during 2014; however, even if pared back, the consequences could still have major financial and operational ramifications for the industry, and will merit discussions at the board level. In this article, we explain the origins of the tax, the features of the tax as currently proposed, and what might happen next. We also highlight some of the specific concerns for members of the board, including the CEO, CFO, CRO and COO. We conclude with some practical advice as to what organizations can be doing now to prepare, recognizing that the scope and timeline for E.U. FTT remains somewhat uncertain.

1 The authors would like to thank Geoff Lloyd and Nigel Nelkon of EY LLP for their helpful comments. The comments made in this article are those of the authors and are in no way representative of the views of EY LLP, its members and partner organizations. All remaining errors are solely the responsibilities of the authors.
As governments around the world seek ways to repair their economies, politicians in a number of countries see taxation of the financial services sector as a politically favorable solution. Not only would it raise much needed revenue but it may be perceived by many as targeting those institutions largely responsible for the crisis. The implications are far reaching for financial institutions, and a cause for concern for all members of the C-suite.

This article examines the so-called financial transaction tax (FTT) and discusses its implications on financial services firms and the steps they need to take in preparation for its arrival.

The concept of financial transaction taxes is not new

Taxing financial transactions is by no means a new concept. Switzerland, for example, had introduced its Securities Transfer Tax (Umsatzabgabe) back in 1917. The tax was later revised in 1971. In 1986, the U.K. introduced Stamp Duty Reserve Tax on cash equities, supplementing its stamp duty charge which has been in place since 1694.

John Maynard Keynes wrote in 1936 that “It is usually agreed that casinos should, in the public interest, be inaccessible and expensive. And perhaps the same is true of the stock exchanges. The introduction of a substantial government transfer tax on all transactions might prove the most serviceable reform available, with a view to mitigating the predominance of speculation over enterprise in the United States.”

In 1972, James Tobin recommended during a series of lectures at Princeton University the introduction of a currency transaction tax. He believed that such a tax would help reduce exchange rate volatility by reducing currency speculation, which was deemed essential in a post-Bretton Woods era, as well as help raise revenues and make national economic policies less vulnerable to external shocks.

From a European Commission perspective, a financial transaction tax is regarded as the preferred way to make the financial services sector pay for the recent and potentially future government bailouts, and of course help raise revenues. They are also hoping that such a tax would deter riskier behaviors, such as high frequency trading.

The E.U. FTT proposal has a further dimension. Generally, taxes have historically remained within the competence of the Member States rather than the E.U. However, the E.U. FTT is now being regarded as an opportunity to demonstrate that an E.U.-wide tax is workable. Hence, the European Commission has stressed that such a tax would help “harmonize” at an E.U. level existing transaction taxes already applied by individual Member States (such as France and Italy).

The first E.U. Commission proposal published in late 2011 required implementation by all Member States. However, this proposal was rejected by those E.U. Member States heavily reliant on financial services, including the U.K., Ireland and Luxembourg. Failure to achieve unanimous agreement amongst Member States was sidestepped by the use of a little-used mechanism called “Enhanced Cooperation.” Under this procedure, nine or more Member States may adopt legislation to be applied by those states, but not by other E.U. Member States. In late 2012, 11 countries (Austria, Belgium, Estonia, France, Germany, Greece, Italy, Portugal, Slovakia, Slovenia and Spain) (the E.U.11), agreed to apply the enhanced cooperation procedure for an E.U. FTT. The introduction of E.U. FTT under enhanced cooperation requires unanimity amongst all the participating Member States; non-participating Member States have no formal right of veto.

Following the agreement amongst the E.U.11, the European Commission published a revised proposal in February 2013, which was based on the initial 2011 proposal but modified to take account of the fact that only 11 Member States would be participating.

Under that proposal, the E.U. FTT would apply not only to equities (as is typical for existing transaction taxes), but also to fixed income instruments and derivatives. A minimum taxation rate would be applied across all of the participating Member States, with the option to apply higher rates by individual countries if they wished to do so.

Apart from the range of instruments potentially in scope, careful inspection of the proposal reveals a number of design principles...
that significantly extend the reach and impact of E.U. FTT beyond the ambit of most transaction taxes currently in force. These include:

- An issuance rule, with taxes becoming due on transactions involving instruments that were issued or treated as issued in one of the 11 Member States (e.g., German bonds, Belgian equities and, it would seem, derivatives traded on a German exchange)
- A residence rule, automatically bringing into scope all transactions entered into by a financial institution established in one of the 11 Member States (e.g., any trade entered into by a French or German bank)
- A deemed residence (or counterparty) rule, where a financial institution enters into a transaction with a counterparty that is established in one of the 11 Member States, irrespective of whether that counterparty is a financial institution (e.g., a trade entered into between a U.K. bank and either a French bank or a French corporate client)
- Taxation of both buy and sell side of in-scope transactions which, when applied in conjunction with the residence rule, effectively brings global financial institutions firmly within the grasp of the tax
- Lack of exemptions for market making, funding and intra-group trades
- Lack of netting and credit mechanisms (unlike VAT, there is a cascade effect where every intermediate step is taxed in full)
- Same day settlement for electronic transactions, with joint and several liability for unpaid tax
- Extensive anti-avoidance rules

If the tax were to be implemented broadly as proposed, it would effectively become a broad tax on global financial institutions, given that most will have some dealings with counterparties and/or instruments that fall within the scope of the tax.

That said, it is widely thought that the architecture of the tax will change significantly between the draft published in February 2013 and the final piece of European legislation and its implementation into national law.

This viewpoint has emerged as a result of a number of factors:

- Legal challenges and uncertainties: the U.K. has brought a formal complaint about the E.U. FTT to the Court of Justice of the European Union (CJEU), alleging that both the issuance rule and, in particular, the counterparty rule are contrary to E.U. law and accepted norms of public international law. The CJEU has not yet considered the merits of the U.K.’s complaint. However, the U.K.’s challenge has been buttressed by an opinion rendered by the E.U. Council Legal Services, who expressed the viewpoint that the counterparty rule was contrary to E.U. law. Although the European Commission has since robustly defended its position, it is fair to say that the position is, at least so far as the counterparty rule is concerned, ambiguous.
- Political deadlock: matters have not moved on materially over the course of 2013 among the participating Member States since the EU Commission’s draft proposal of February 2013. Reports from meetings involving only the participating Member States and those involving all Member States suggest that there has been significant disagreement among the participating Member States and that, generally, there has been a reluctance to drive through the proposal in the form proposed by the European Commission. Notably, France, which was a prime mover behind the enhanced cooperation process, is understood to have moved its ground, now throwing its weight behind a more narrowly focused equities-only tax rather than a tax applying to all instruments.
- Significant advocacy efforts: significant advocacy efforts have been undertaken opposing the tax – not just by global financial institutions, but also by end investors, such as pension funds, and participants in the non-financial sector, who would also likely bear much of the economic burden of an E.U. FTT.

As a result of these factors, the original timetable envisaged in the February 2013 proposal for the tax to be introduced on 1 January 2014 has been missed. The time it will take for participating Member States to reach an agreement and the process to transpose final European legislation into local law, as well as some allowances made for the time it will take FS organizations to implement the change suggests that a later date for implementation, in 2015 or even 2016, is more likely.

---

4 It is possible that a “work in progress” draft will be published in Q1 or Q2 of 2014, with the expectation that a number of changes and exemptions will be included.
The precise scope and shape of an E.U. FTT also remains unclear. Much will depend on progress made in the political discussions among the participating Member States.

- **Outcome of the German elections**: in late 2013, Angela Merkel’s CDU agreed to form a “grand coalition” with the Social Democrats (SPD), who have been enthusiastic supporters of the E.U. FTT concept. The coalition agreement between the CDU and SPD expresses support for an E.U. FTT based on the Commission’s draft of February 2013, subject to certain caveats. Following the election, many market commentators expect Germany to provide a stronger lead with respect to finalizing the tax.

- **Presidency of the European Council of Ministers**: the Presidency of the European Council will sit with two participating Member States throughout 2014: Greece, from 1 January to 30 June 2014 and Italy from 1 July to 31 December 2014, which may influence the pace at which E.U. FTT is debated and decisions reached.

Against that backdrop, the prudent approach would be to prepare for the introduction of an E.U. FTT, albeit that the timing and scope of such a tax still remain unclear. At the time of writing, it has been reported that the European Commission has expressed the hope that the participating Member States could reach agreement by the summer of 2014, with a compromise being facilitated through a “staggered” introduction of the tax.

**CEOs will be concerned by the potential impact on profitability, and thus shareholder returns**

There have been many estimates about the amount of revenue that E.U. FTT will generate on behalf of national treasuries at the expense of financial institutions. These range from the “official” Commission estimates of €34b per annum (after behavioral changes resulting from the tax), up to an eye watering €170b per annum (not taking account of likely behavioral changes).^5^ That said, the orders of magnitude are sufficient to place the issue on the CEO radar.

The primary concern will relate to the financial viability of lines of business, and competitive positioning. While there is arguably no such thing as a level playing field in the world of capital markets, E.U. FTT is likely to create overnight competitive advantages and disadvantages. At a minimum, it will require broker-dealers to revisit their strategies for both operating in the affected European countries and also their approach to trading impacted products.

If the counterparty rule remains in the final legislation (although we believe that there is significant pressure from key Member States, such as France, Italy and Spain, to remove altogether any sort of residence rule and move to an issuance-only basis of taxation), those institutions whose business is booked in legal entities outside of the participating Member States (principally London, Zurich, New York, Chicago, Hong Kong, Singapore and Tokyo) will be better placed to capture business from those established within the zone. It is conceivable that we may see subsidiaries take the place of branches of E.U. 11 domiciled legal entities.

For asset managers, in addition to shifting their businesses to counterparties outside the participating Member States, it is possible we may see, for example, some relocation of collective investment schemes away from the participating Member States as well as strategic changes to how assets of pension funds resident in the E.U. 11 (assuming not exempt from the E.U. FTT) are managed. Indeed, market reports indicate that there may have been a certain rebalancing of portfolios by fund managers following the introduction of FTTs in France and Italy; it remains to be seen whether a similar pattern will emerge if E.U. FTT is introduced. There may, of course, be other important structural changes by asset managers, depending on how dealings in fund units/shares and investments by funds are to be impacted by the Directive.

Those financial institutions disadvantaged in this way may then face the difficult decision of deciding whether to change their legal entity structures/domiciles and booking models. CEOs will need to weigh up the commercial reasons for doing so, versus the cost and any political ramifications of such a course of action.

---

The CFO will be concerned about the implications for funding
Throughout 2013, we met with or spoke to more than 100 financial institutions globally to discuss the impacts of E.U. FTT. Their single biggest commercial concern was the potential impact on overnight funding through the taxation of repo transactions; this was confirmed during our industry survey in Q3 2013. ICMA’s analysis suggests that, assuming FTT was levied on both ends of a repo transaction at 0.10%, a repo market-maker would have to charge a spread on an overnight repo of 7,205 basis points. It goes on to suggest that the nearest horizon at which the tax might be tolerable would be greater than one year, and that even if repo activity beyond 6 months survived, the market would shrink by at least 66%. Given the importance of overnight repos to financial institutions’ funding strategies, organizations will have to find alternative approaches. ICMA goes on to suggest that organizations may instead switch to unsecured deposits outside the E.U. 11, which in turn results in capital flight, a reduction in lending to the real economy and undermining of the single market.

It is clear that the full consequences of the lack of an exemption for funding repos were not fully considered by the Commission in making its legislative proposals. We have observed significant lobbying activity in this respect. Central Banks are believed to be concerned about this matter and, to this extent, our expectation at the time of writing is that the final tax is likely to provide exemptions or some other relief for repos used for funding purposes. Indeed, the Commission itself has already raised the possibility that the applicable tax rate for repos could be reduced from 10 bps to 1 bp.

A second concern for the CFO is the potential lack of an exemption for intra-group transactions. This may be a driver for legal entity rationalization.

The CRO will be interested to learn the impact for risk management and liquidity
The CRO will have four primary concerns with respect to EU FTT.

First, the tax applies to transactions regardless of their purpose. There is currently no planned exemption for hedging transactions (although the European Parliament, and indeed the French Finance Minister, has suggested that such a distinction be introduced); hence the cost of managing risk rises. Given that the minimum proposed rate of tax on derivatives is 1 bp as opposed to 10 bps for equities, financial institutions may look to hedge positions via derivatives rather than physicals. This may mean less perfect hedges and increased risk. Furthermore, banks may look to hedge positions with instruments that are outside the scope of FTT altogether.

Second, there is the impact on liquidity and potential market dislocation. Volumes in certain products and markets will be significantly reduced. Although, it is pretty much impossible to model the impact on volumes with any degree of certainty, the European Commission estimates that derivative volumes could fall by 70%, though other market commentators have suggested that a figure of 50% is more likely. The European Commission also expects volumes of bond and equity trades to fall by 15%. Interestingly, the reduction in volumes of French equities in 2012, following the introduction of French FTT over cash equities, was estimated by TABB group to be 26%. In May 2013, Per Callesen, a National Bank of Denmark governor, suggested that FTT is a “tax on market liquidity”, and conflicts with the global regulation to increase banks’ liquidity buffers.

Third, there will be an increase in operational risk. The scale and complexity of the proposed tax, combined with the number of organizations impacted across each part of the financial services ecosystem – from banks to broker-dealers, financial market infrastructure providers, asset managers and pension funds – has caused many to worry about the industry’s ability to implement the proposal in a safe and orderly manner.

Fourth, the potential increase in counterparty credit risk cannot be ignored given the existence of joint-and-several liability. Where both parties to a transaction are jointly and severally liable, and one party defaults on the tax liability, the tax authorities may be able to pursue the other for payment, albeit that this risk is arguably more likely where the party in default is outside the E.U.
COOs and heads of IT will play a major role in leading the organization’s response

Some of the potential responses to E.U. FTT as described above will need to be led by the COO.

We would expect some organizations to change their booking models, routing business away from legal entities established in the E.U. 11 countries. Changes to legal entity structures will also need to be addressed by the COO from an operational and efficiency perspective.

Operational compliance with the E.U. FTT will be a step-change from any previous operational tax, including FATCA. The calculation, collection and payment of E.U. FTT will need to be integrated into the front to back product trade flows. As it is currently proposed, organizations will be challenged to assess the impact from execution onward, build a suitable collection/payment platform and then resource the ongoing process from already stretched operations resources. Against a backdrop of profit margin compression, other regulatory, risk and tax obligations and a commitment to automated STP (straight through processing), meeting the demands of E.U. FTT will not be an insignificant undertaking.

A further complication arising from E.U. FTT is in terms of trade execution, market making prices and spreads impacting front office algorithms and trading models. Clients will also need transparency on their execution data in order to amend their own trading strategies.

Most organizations solved for French and Italian FTT tactically but the features of E.U. FTT, notably overall volumes, product complexity and same day settlement for electronic transactions demand a more strategic response, founded on a scalable and flexible architecture that can cope not only with any variables arising from the fact that an E.U. level tax must be transposed by 11 different Member States (each with their own legal systems), but also with other transaction taxes in the event that other Member States sign up to the tax or, indeed, that other non-E.U. jurisdictions decide to follow suit and introduce their own transaction taxes.

Another significant concern for the European Commission, participating Member States, and market participants is that the mechanism for collecting and reporting that the tax has not yet been clearly defined. However, unwelcome the proposed tax is for financial institutions, the lessons learnt from French and Italian FTT experience suggests that an onerous or ambiguous collection reporting and collection mechanism will compound the problem.

Technologists will be able to build or buy sophisticated rules engines. However, the complexity of the rules and inherent challenges with respect to the quality and availability of transaction and reference data could result in numerous transactions needing to be investigated manually, thereby introducing further costs and delay to the trade lifecycle.

Chief Information Officers and their teams will quickly learn that E.U. FTT represents a substantial book of work in its own right. Given the asset classes in scope and the global reach of the tax, the number of source systems, data repositories and operational systems that will need to be reviewed and enhanced to cope with the tax will be significant. Clarity of rules and an implementation timetable that allows for a planned and structured solution to be established are both required.

Defining the technology solutions to address the requirements of E.U. FTT will touch on many areas already being addressed through other initiatives, and operations and technology teams will need to explore the commonalities between E.U. FTT and other regulatory initiatives (such as EMIR, MiFID II, KYC/AML, FATCA) that demand changes to trade reporting, transaction reporting, client onboarding and reference data.

Non-executive directors will play a key role as “the conscience of the Board”

Non-executive directors will observe the developments of FTT and advise the board in their independent and impartial capacity. Non-executive directors are often more attuned to matters such as reputational risk and external perception, and will legitimately challenge any responses to FTT that could have adverse reputational consequences.

What happens next?

The next major milestone will be the reaching of an agreement in respect of the architecture, scope and rates of tax amongst the 11 Member States. This will be an intensive process, as different
Member States have different agendas, and have different views and approaches to the impacts of the tax.

Certain countries, including France, Italy and Spain, are thought to be keener on restricting the tax to an issuance rule only, and are also supportive of removing bonds from the scope. As noted above, the German position is expected to be clarified, and be more obviously supportive of some form of E.U. FTT, following the recent election and formation of a “grand coalition” between the CDU and the SPD. Other countries, however, fear that changes in the scope or the reach of the tax may materially reduce the amount of revenue that they hope to collect.

Assuming a political consensus is reached, and the tax is then passed into E.U. law, it would then be necessary to transpose the E.U. rules into national law in each of the participating Member States. This process is naturally constrained by the shape of the legislation agreed in Brussels, although within the overall agreed framework there will still be considerable scope for local differences in terms of definitions, interpretation, and any parts of the law that need to be adjusted to fit with local jurisdictions.

A larger challenge will be how the enforcement and collection processes work, not just across the 11 Member States but globally. Little thought had been afforded to this during the initial work on the tax, and much remains unclear.

There are therefore a number of fundamental questions where detailed answers are yet to be provided.

**Reporting and settlement of the tax**

The reporting and settlement of the tax will be heavily dependent on the final scope of the tax. In the event that the tax ends up being “issuance only”, with the residence rule removed, it would be reasonable to expect that the market infrastructure participants (exchanges, clearing houses, CSDs) in each of the 11 countries would likely play a key role in tax collection and reporting (in the same way that U.K. SDRT is collected through the CREST system by Euroclear; the Central Securities Depository).

If, however, the residence rule remains in the final version of the tax, it is far from clear how the tax would be reported and settled.

While the proposal sets out a hierarchy as to which country is entitled to tax revenues from particular transactions, it remains largely silent on collection mechanisms. Financial institutions could be faced with the added complexity of reporting and settling to 11 different jurisdictions, each of whom may stipulate their own requirements.

Further grey areas exist around instruments that may fall within the scope of the tax, but where place of issuance would be unclear; for example, how is the place of issuance to be determined for an OTC derivative transaction?

**A question of enforceability**

There is considerable debate on the enforceability of an E.U. FTT outside the 11 participating Member States, and in particular, outside the E.U. itself. In July 2013, two U.S. politicians (Senator Pat Roberts and Representative Tom Price) introduced legislation in the Senate and the House of Representatives to block foreign governments from collecting taxes on securities transactions.

In the same way that certain countries’ laws and jurisdictions prevent compliance with other tax legislation such as FATCA, we can anticipate similar complications in the extra-territorial enforcement of E.U. FTT. Non-E.U. 11 financial institutions in particular could potentially find themselves in an awkward position. They may, albeit reluctantly, feel obliged to pay their way under E.U. FTT, yet conceivably find themselves in a position where it may even be illegal in their home jurisdiction for them to comply.

To this extent, we expect enforceability to become possible only with international agreements with non-EU countries and non-participating Member States.

Notwithstanding the above, and whatever basis of taxation remains in scope, the joint and several liability feature of the tax architecture will be put to the test. While, say, a tier-1 bank in a major economy may agree to pay the tax, its smaller counterparty in an emerging market may choose not to. E.U. FTT, therefore, introduces a new type of risk; “tax credit risk.” We would expect this to become a feature of client onboarding and KYC processes in the future — the question needs to be asked, “will my counterparty pay its share of the tax due on our trades?”
What should financial institutions be doing now?
The pace at which organizations need to start addressing
the challenges of E.U. FTT depends very much on the
implementation schedule and the propensity for further
changes in the proposed tax.

The vast majority of financial institutions are already heavily
weighed down by a commitment to deliver against a number of
different mandatory regulatory initiatives. We would not expect
organizations to defocus on more pressing deadlines in favor of
E.U. FTT.

We are however observing the following activities in the
marketplace:

Determine the impact
Firms are undertaking impact assessments. These vary
considerably from those that are purely commercial, to those that
follow the logic through to potential changes in booking models,
legal entity reconfiguration and funding models.

The booking model and legal entity question is a challenging one,
as there are many drivers for change, and the specter of E.U. FTT
may not be the largest driver.

Firms will need to ensure their use of capital is optimized. At
the same time, many will have views on where they wish to be
regulated, and this will play a role in determining any strategy
that is founded on subsidiarization. Bank separation and “too-big
to-fail” agendas will also be relevant.

Raise internal awareness
Most firms are raising internal awareness. In most cases, the
initial interest sparked in finance or tax. However, the potential
commercial impact has alerted front office teams to the scale of
the issue. Operations, risk management, technology, legal and
investor relations functions are usually engaged to some extent.

Organize the response
Governance structures will vary from organization to
organization. There is no single right answer. However, our view
is that the governance structure should include all relevant
stakeholders, including inter alia heads of desks, middle office,
operations, risk, finance, tax, legal, compliance, technology and
public relations. For organizations with multiple business lines
(e.g., those with broker-dealer and investment management
arms), it is necessary to determine the appropriate level of
connectivity required internally.

It will be essential to determine the levels of delegated
responsibility with respect to decision making, and the escalation
process for making difficult decisions.

We expect that once details of the tax are agreed and ready to be
transposed into law, implementation timescales are likely to be
challenging (in order to expedite the raising of funds via the tax).
Establishing momentum and appropriate cadence will be key.

Finally, we advise financial institutions to stay close to industry
commentators and thought leaders to keep abreast of the latest
developments as they happen. We expect many twists and turns in
the E.U. FTT story as it unfolds, and staying well informed will help
to make the challenging delivery journey as smooth as it can be.

Conclusion
At the time of writing, there is still considerable uncertainty as
to how and when the E.U. FTT proposal is likely to be resolved.
However, our point of view is that the political will behind the
proposed tax should not be underestimated; and therefore it
would be prudent for the market to assume that overall some
form of an E.U. FTT is likely to happen in the next two years.

The tax is likely to merit board level discussion within financial
institutions. The nature of the tax means, and its commercial,
risk management, finance and operational implications, suggest
that most members of the C-suite will have a role to play in
determining a financial institution's response.
Part 2: Tactical

Measuring capital adequacy: supervisory stress-tests in a Basel world

Externalities and macroprudential policy

Directors’ and officers’ insurance and shareholder protection

A long-term approach to Italian banks’ profitability: paradise lost?

The introduction of market-based pricing in corporate lending

Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

Valuation effects of termination of cross-listings
Measuring capital adequacy: supervisory stress-tests in a Basel world

Larry D. Wall
Executive Director, Center for Financial Innovation and Stability (CenFIS), Research Department, Federal Reserve Bank of Atlanta

Abstract
The U.S. is now committed to using two relatively sophisticated approaches to measuring capital adequacy: Basel III and stress-tests. This paper shows how stress testing could mitigate weaknesses in the way Basel III measures credit and interest rate risk, the way it measures bank capital and the way it creates countercyclical capital buffers. However, the paper also emphasizes the extent to which stress-tests add value will depend upon the exercise of supervisor discretion in the design of stress scenarios. Whether supervisors will use this discretion more effectively than they have used other tools in the past remains to be seen.

1 I would like to thank Scott Frame, Gillian Garcia and Bev Hirtle for helpful comments on a prior version called “Supervisory stress testing: will the long-run benefits exceed the on-going stress.” Any remaining errors are solely the responsibility of the author. The opinions expressed in this paper are those of the authors and do not necessarily reflect the view of the Federal Reserve Bank of Atlanta or the Federal Reserve System.
1. Introduction

The U.S. is now committed to using two relatively sophisticated approaches to measuring capital adequacy. The first approach, stress-testing, was pressed into service in early 2009 to help restore confidence to the banking system and was made into an ongoing annual requirement by the Dodd–Frank Act. The second measure, Basel III, was adopted as a regulation by all three federal bank supervisory agencies in July 2013 for the implementation to start on 1 January 2014. Both of these measures are not only more sophisticated, but also more complex, costly and time consuming to implement than the capital adequacy measures brought in place in the U.S. in 2009.

Capital adequacy ratios have been an important tool of prudential supervision dating back prior to the adoption of the first Basel Accords in 1988. Conversely, stress-testing has a shorter history and generally played less of a role in micro-prudential supervision. Prior to 2009, stress-tests were used to analyze the overall financial stability of the banking system and the risk in selected parts of banks’ balance sheets.

The Supervisory Capital Assessment Program (SCAP) in 2009 took these stress-tests to a different level and that change was philosophical, with the blending of microprudential approaches to measuring capital adequacy. The first approach, Basel III, was adopted as a regulation by all three federal bank supervisory agencies in July 2013 for the implementation to start on 1 January 2014. Both of these measures are not only more sophisticated, but also more complex, costly and time consuming to implement than the capital adequacy measures brought in place in the U.S. in 2009.

Capital adequacy ratios have been an important tool of prudential supervision dating back prior to the adoption of the first Basel Accords in 1988. Conversely, stress-testing has a shorter history and generally played less of a role in micro-prudential supervision. Prior to 2009, stress-tests were used to analyze the overall financial stability of the banking system and the risk in selected parts of banks’ balance sheets.

The Supervisory Capital Assessment Program (SCAP) in 2009 took these stress-tests to a different level and that change has continued in the U.S. under the follow-up Comprehensive Capital Analysis and Review (CCAR) and the Dodd–Frank Act Stress Tests (DFAST) programs. An important part of the change was philosophical, with the blending of microprudential and macroprudential supervision as emphasized by Hirtle et al. (2009). However, the operational changes were also substantial according to Bookstaber et al. (2013), who call the current version of stress-testing “Stress Testing 2.0.” They observe that starting with SCAP, in the U.S., there was a substantial expansion of the scale and granularity of the data collection and modeling used in stress-testing.

Given the existence of one costly risk-based measure, an obvious question is what is the incremental value of implementing a second costly measure? The purpose of this study is to provide an answer to this question from the perspective of how supervisory stress-tests could mitigate problems with Basel III. The primary reasons for asking how the stress-tests can add value is that, as we shall see, the supervisors have considerably more flexibility in the implementation of the stress-tests and they can use that flexibility to minimize the impact of Basel III’s weakness. A secondary reason is that Basel III purports to measure the full range of bank risks whereas the stress-tests only measure the losses associated with a handful of specific scenarios.

The paper is organized as follows. Section 2 provides an overview of the workings of Basel III capital ratios (Pillar I) and the stress-tests as conducted in the U.S. Sections 3, 4 and 5 discuss the potential contributions of stress-testing to overcome weaknesses in Basel III’s approach to measuring credit risk, interest rate risk and bank capital. The sixth section considers how stress-testing could provide an alternative method of implementing countercyclical capital buffers that may be less subject to political pressure than the mechanism in Basel III. The paper concludes with a summary of the possible contributions of stress-testing to the measurement of individual bank capital.

2 Geithner (2009) announced the stress-tests and the Board of Governors of the Federal Reserve System [hereafter BOGFRS (2009)] announced the results. Wall (2013b) provides an overview of the history of stress-tests used to measure capital adequacy in the U.S. and E.U.
4 The U.S. had not adopted Basel II at the start of 2009. Instead, the U.S. was operating under a combination Basel I (including its amendments) and a leverage ratio that had gross total assets in the denominator.
5 A comprehensive review of this literature on capital adequacy ratios would go far beyond the scope of this paper. Wall (2013b) provides a condensed history of the capital regulation focusing on the development of the Basel capital accords.
6 Schuermann (2012) provides a short review of the literature on stress-testing. See Borio et al. (forthcoming) for a critical review of the use of stress-tests for macroprudential purposes.
7 Wall (2013b) provides a brief overview of stress-testing for capital adequacy purposes, including SCAP, CCAR and DFAST. See also Bernanke (2013) for a discussion of Federal Reserve stress-testing, starting with the SCAP. See also http://www.federalreserve.gov/bankinforeg/stress-tests/dodd-frank-act-stress-testing.htm [accessed 21 October 2013] for a discussion of the similarities and differences between CCAR and DFAST.
8 Another motivation for stress-testing conducted using the expanded scale and granularity of the U.S. tests would be to provide more sensitive tests of financial stability. Bookstaber et al. (2013) provide both a set of relatively minor variations on current stress-testing that they call “Stress Testing 2.1” and bigger set of changes that they call “Stress Testing 3.0”, which would make the exercises more useful for financial stability purposes.
9 In principle, it would be possible to use stress-tests to estimate the full distribution of losses associated with a bank’s portfolio. However, this would require supervisors to adopt one or more macroeconomic models, simulate these models thousands of times and use each of these simulations as the scenario for the stress-tests applied to each bank. Such an application of stress-tests appears unlikely for the foreseeable future, given that thus far the individual bank tests have been run on at most a handful of scenarios.
2. A comparison of Basel III and stress-testing

Basel III and stress-testing, as implemented in the U.S., both rely on projections of losses in an extreme scenario to evaluate the adequacy of individual bank’s capital. Both measures require the estimation of statistical models. In many cases, banks and the supervisors have data covering at most the past two decades, a period that includes only one severe business cycle.

However, the two measures also differ in some fundamental ways according to Wall (2013b). Basel III provides an unconditional static measure, with the risk adjustment occurring in the risk weighting of assets (the denominator of the capital adequacy ratio). In contrast, the stress-tests, as currently applied to measuring individual bank capital adequacy, are conditional, dynamic measures, with the risk adjustment occurring via reductions in capital (the numerator of the capital adequacy ratio). Basel III is an unconditional, static measure in that it measures capital adequacy at a single point in time using a process that does not depend upon projections of the future state of the economy. That is, the calculation of the Basel III ratios uses data on past performance to estimate the loss distributions associated with various portfolio positions. The estimated loss distributions are then used to calculate the expected losses in the extreme tail of the distribution. The current portfolio positions (assets and derivatives) are then summed using weights derived from the expected losses to calculate the denominator of the Basel III ratios. In effect, Basel III derives a generic, severely adverse scenario for each portfolio category from that category’s own (recent) past experience.

Basel III provides three different methods of risk weighting assets. In the standardized approach, the supervisors provide the risk weights to the banks based on their estimates of the riskiness of different assets. In the internal ratings based (IRB) approach, banks supply their own estimates of the probability of default, which is then entered into supervisory models to obtain risk weights. Finally, in the “advanced IRB” approach, banks also come up with their own estimates of loss given default and exposure at default, along with their estimate of the probability of default.

In contrast to the relatively sophisticated approach to calculating risk-weighted assets for the denominator of the Basel III ratio, the calculation of the measure of capital is mechanical. Basel III uses three different measures of capital: common equity tier-1, which includes items, such as common equity and retained earnings that are available to absorb losses on a going-concern basis; tier-1 capital, which includes other perpetual instruments that are subordinated to the deposits and subordinated debt of the bank and that meet additional criteria; and tier-2 capital, which includes items available to absorb losses only on a gone-concern (in resolution) basis, such as subordinated debt. Basel III also provides for certain mandatory deductions from capital, such as subtracting the value of certain intangible assets from tier-1 capital. However, the values of each of the items in capital and the deductions from capital are taken from the financial statements of each bank holding company (BHC).

The stress-tests begin with several different regulatory measures of capital adequacy. Among these measures are the Basel III risk-based ratios. However, the stress-tests are dynamic in that they simulate how these regulatory ratios would evolve over time and are conditional in that the results are calculated for a specific scenario for the economy. The primary focus of the stress-tests has been on estimating changes in accounting capital following the standards set by the Financial Accounting Standards Board (FASB) in the U.S. or the International Financial Reporting Standards (IFSR) in the E.U. The estimates of the change in accounting capital are based in part on estimates of each bank’s losses in each portfolio in each period. However, given that the stress-tests are dynamic, they also include estimates of each bank’s pre-provision net revenue (PPNR).

The first step in conducting a stress-test is to estimate the historical impact of economic variables, such as GDP growth on bank’s losses and PPNR, given certain important characteristics of each bank’s portfolio. The next step is to develop one or more internally consistent scenarios for the future evolution of the economy. Estimates of losses and PPNR for each period in each scenario are then obtained by plugging in the characteristics of the bank’s portfolio and the stress scenario into the bank’s and the supervisor’s models estimated using historical data. The projected losses and projected capital distribution are subtracted
from PPNR to estimate each period’s change in capital. The capital at the end of each period is then its value at the start of the next period.

One way of illustrating the differences between Basel III and stress-tests is to imagine them as different approaches to looking for risk. Basel III casts a dim light over a wide range of possible scenarios. Basel III’s use of historical loss distribution data allows it to predict losses in tail of the distribution across a wide variety of scenarios. However, Basel III cannot say very much about what may happen in any particular scenario. In contrast, each individual stress-test casts a very bright light but only on one particular scenario. To the extent the models are correct, the stress-test is intended to provide a good estimate of what happens in a particular scenario. The results of the test of any one scenario are likely to be representative of the losses that would arise in similar scenarios, but such a result is not guaranteed. Moreover, there is no reason to expect that any given scenario will be predictive of the results of a very different stress scenario.

3. Credit risk mis-measurement
Arguably, the biggest risk to banks is that of credit risk. This section considers potential weaknesses in Basel III’s measurement of credit risk and how stress-tests may mitigate some of these problems.

In order to understand the vulnerability of Basel III or any other measure of capital adequacy, it is helpful to first understand how banks will respond when confronted with a possible deficiency in capital. In general, a bank approaching deficiency can implement some combination of two possibilities: increase the denominator (retain more earnings or issue new capital or both) or reduce the numerator (report taking less risk). Although bank supervisors would often prefer that banks retain more earnings and/or issue new capital, banks generally view equity capital as their most costly source of funding. In many cases, banks find that reducing their reported risk would be less costly. A bank that seeks to reduce its reported risk may employ some combination of the following three options: (a) shrink its overall portfolio while retaining the same allocations within the portfolio, (b) reallocate part of its portfolio away from positions with higher capital charges and toward positions with lower capital charges and (c) change the way risk is measured so that, holding the portfolio constant, the estimated risk exposure is lower than in the original model.

A bank that shrinks its overall portfolio while retaining the same allocations will reduce its risk level. Supervisors’ concern here is that if all banks shrink their portfolios at the same time, the resulting reduction in the supply of credit may reduce economic growth (or deepen a recession), which in turn would produce an offsetting increase in the risk of banks. The Basel III requirements provide no mechanism for preventing banks from meeting their requirements solely by shrinking their portfolios. On the other hand, stress-tests may be designed to force banks to issue capital as the U.S. did with the 2009 SCAP. A policy of forcing banks to issue new capital will not necessarily result in more new loans, but it removes inadequate capital as a reason for banks to refuse to make new loans or extend existing loans.

The next subsection considers how banks reallocate their portfolios when subject to binding capital regulation. The following two subsections examine the potential for systematic errors due to limited data and biased measurement approaches.

3.1 Portfolio reallocation
Portfolio reallocation could be an effective way of increasing capital adequacy if the risk weights in Basel III are the “correct” weights. Portfolio reallocations would be in the direction of reducing exposures to positions where the bank had previously underestimated the asset’s risk and possibly increasing positions where the bank had overestimated the risk.

However, there are a variety of reasons to expect that the Basel III contains many random errors. The standardized approach is intended to provide a rough approximation of the average

---

12 For example, the results of a severe economic downturn scenario are unlikely to be very informative about the consequences of a scenario with sharp increases in inflation and interest rates.
13 See also Wall and Peterson (1996) for a discussion of banks’ response to binding capital requirements.
14 While the stress-tests can be designed to force banks to issue new capital, even the stress-test approach would have difficulty forcing banks to make new loans to support economic growth. Supervisors are (quite correctly) concerned about forcing banks to make new credit risky loans that the banks judge to have too low of an expected return for their risk.
risk of various risk buckets across banks, implying that it will measure the risk of individual banks with error unless these risk buckets have exactly the same risk within all banks. The IRB approaches are more sensitive to each bank’s position. However, Haldane (2011) estimates that Basel III has more than 200,000 risk buckets. The weights for each of these buckets must be estimated using complex statistical procedures, procedures which themselves recognize the likelihood of estimation errors.

The combination of unavoidable estimation errors and banks’ portfolio reallocations to reduce their capital requirements implies that actual banks’ capital adequacy will be less than the stated standards of Basel III. In order to illustrate this problem, start by assuming that Basel III is correct on average. That is, if Basel III was applied to existing portfolio, those cases where the risk weights were too low would be almost exactly offset by cases where the weights are too high. However, banks will respond to binding regulations by seeking to reduce their Basel III capital requirements in the way that has the least adverse impact on shareholder wealth. Given the existence of the safety net (deposit insurance, lender of last resort, “too-big-to-fail” implicit subsidies), this implies that the bank portfolios should show the largest reductions in those areas where the Basel III risk weightings are too high relative to actual risk and the smallest reductions in those areas where the risk weightings are too low. The result of this portfolio reallocation is that even if the Basel III standards were correct on average prior to banks reallocating their portfolios, total risk-weighted assets (RWA) will be too low after banks finish reallocating their portfolios.

The measurement errors associated with stress-tests are not necessarily smaller than those with Basel III but they are likely different. Basel III uses historical data to estimate unconditional loss distributions, whereas stress-tests use historical data to estimate the relationship between losses and specified economic variables (such as GDP growth). The difference in estimation methods implies that the errors from Basel III are unlikely to be perfectly correlated with those from the stress tests. For some of the asset classes where Basel III underestimated the losses, the stress-test loss estimates will overestimate the losses or at least underestimate by a smaller margin. Thus, it is likely that stress-tests will reduce banks’ incentives to shift their portfolios toward assets whose RWA would be too low under Basel III.

The stress-tests also provide supervisors with the option of designing stress scenarios that discourage investment in assets that the supervisors think have too low of a risk weight under Basel III. For example, suppose that the supervisors conclude that the Basel III weights underestimate the risk of commercial real estate lending and as a result they see banks increasing their portfolios allocations to these loans. The supervisors could require banks to run a stress-test that has an especially adverse scenario for commercial real estate lending.

3.2 Biased credit risk measures
The problem of banks shifting their portfolios toward assets underweighted by Basel III is compounded if the source of the underweighting is due to something other than a random error. Basel III has at least two weak links in its estimation of credit that could lead to systematic underestimation of the riskiness of many assets: some important risky scenarios are not observed in the relatively recent historical data and banks may use their discretion to pick models that underestimate risk.

3.2.1 Bias due to data
The loss distribution models in Basel III have to be estimated from historical loss distributions. If the historical data is drawn from a benign economic period with relatively low losses, the estimated loss distributions may understate the true potential losses during a more volatile period. For example, estimates of potential losses on residential real estate obtained during the early 2000s, when few loans defaulted for a loss because of the general trend toward higher house prices, would not necessarily reflect the potential for losses if national average housing prices drop.

A possible solution, if the recent losses have been low, would be to go further back in time. The problem is that the further one goes back in time, the more likely they are to incorporate structural change(s). For example, the last time house prices fell nationally was during the 1930s. As Rowe (2013) points out, such structural changes undermine the assumption implicit in most statistical analysis that the underlying stochastic process is stable. For example, the economy and housing finance systems of the early 2000s were very different from those of

15. The following subsection will discuss a variety of reasons why the Basel III weights might on average be too low.
the 1930s, implying that estimates derived from 1930s data may not be very predictive of the losses that did happen in 2000s after house prices started to fall.

The loss models used in stress-tests are also estimated using historical data. However, stress-tests provide supervisors with an additional degree of freedom to pick stress scenarios that are not observed in the recent data. How stressful the scenario should be would depend upon the supervisors’ beliefs about the stress-test loss models. If the supervisors’ believe the loss models would deliver reasonably accurate estimates of the likely losses, they only need to specify a scenario that incorporates their concerns. For example, even if the data used to estimate the models lacks a 10% national housing decline (as would have been the case prior to 2006), the supervisors could require banks to calculate their expected capital positions given a 10% decline.16

A bigger problem arises if the supervisors believe their models would underestimate losses in a stress scenario. This problem could arise, for example, if the historical data suggests a linear relationship between losses and their scenario but the supervisors believe the true relationship is convex (an increase in the severity of conditions leads to a more than proportional increase in losses). For example, suppose the model predicts that a 20% housing decline would lead to 4% losses, whereas the supervisors judge the losses would be substantially more than 4%. Even with this modeling problem, the supervisors could still design stress-tests that would reveal appropriately large losses by requiring banks to use an even more severe scenario. However, in order for this to work, the supervisors might need to require banks to hold capital to withstand an extraordinarily severe scenario that almost all knowledgeable observers would argue is not plausible. Although it is easy to argue that in principle supervisors should impose implausible scenarios when they think their models significantly underestimate likely losses, doing so in the face of bankers’ and borrowers’ objections is likely to prove challenging in practice.

3.2.2 Bias due to bank modeling

Banks estimate the models to determine the Basel III risk weights under both the IRB and A-IRB approaches. As a result, wherever there is any discretion to be exercised, it is the banks that select the data and statistical procedures. This raises an obvious concern that banks will pick the data and procedures that produce the lowest risk weights [Blum (2008)]. This concern is somewhat reduced by banks’ need to obtain prior supervisory approval for their modeling. However, it is not practical for the supervisors to review all of the decisions made in the estimation of more than 200,000 risk weights in Basel III model. In order to better understand the extent of cross-sectional dispersion in risk modeling, the Basel Committee on Banking Supervision (2013, p. 7) asked a group of large banks in Europe, North America and Asia to evaluate a common set of “largely low-default” credits. Given a 10% capital benchmark, the study found that the risk-weights calculated by the banks could vary by as much as 1.5 to 2 percentage points in either direction, albeit most were within 1 percentage point.17

Stress-tests are also based in part on models estimated by the banks. However, unlike Basel III calculations, the U.S. stress-tests are based, in part, on supervisory models. That is, the Federal Reserve estimates its own models and then applies those models to detailed information supplied by banks about their individual portfolios. As a result, the supervisors can compare their estimates of losses and PPNR in each scenario with the estimates produced by each of the banks subject to stress-testing. Additionally, the supervisory stress-tests are conducted at the same time using the same set of scenarios, which facilitate the comparison of estimated losses and PPNR across banks. This puts supervisors in a better position to identify bank models that produce significantly lower loss estimates. Further, because the U.S. supervisory stress-tests are conducted as a part of an overall review of banks’ capital planning, material weaknesses in a bank’s stress-test modeling

16 The last time national house prices declined was the Great Depression, according to the United States Council of Economic Advisors (2012, p. 101).

17 Cross-sectional comparisons of Basel III weights have been used by the Danish supervisory agency Finanstilsynet, according to Newton (2013). However, reliance on purely cross-sectional comparisons may recreate a problem similar to that observed with the U.S. supervision of capital adequacy prior to the U.S. adoption of numeric standards in 1981. The problem according to Marcus (1983) was that absent firm supervisory standards, although the supervisors could stop individual banks from decreasing their capital ratios to levels notably below those of their peers, they could not prevent the banking industry from gradually reducing its capital adequacy ratios.
remained the responsibility of national authorities. While the Basel interest rate risk in the rest of a bank’s portfolio, its “banking book,” marked-to-market on a daily basis. The provision of capital for the held to profit from short-term price movements and which a bank for the trading book. That is only for instruments that a bank the agreed-upon method of calculating capital adequacy but only [BCBS (1996)] incorporated a measure of interest rate risk into requirements (if any) for interest rate risk were the responsibility exclusively on credit risk. The measurement and associated capital Committee on Banking Supervision (1988) agreement focused for assets likely to be held to maturity (the banking book). Currently, there is no interest rate risk measure measurement because Basel III only incorporates the effect of interest rates on market values for those assets held for trading (the trading book). Currently, there is no interest rate risk measure for assets likely to be held to maturity (the banking book). The first common capital adequacy framework set out by the Basel Committee on Banking Supervision (1988) agreement focused exclusively on credit risk. The measurement and associated capital requirements (if any) for interest rate risk were the responsibility of national authorities. The market risk amendment to Basel I [BCBS (1996)] incorporated a measure of interest rate risk into the agreed-upon method of calculating capital adequacy but only for the trading book. That is only for instruments that a bank held to profit from short-term price movements and which a bank marked-to-market on a daily basis. The provision of capital for the interest rate risk in the rest of a bank’s portfolio, its “banking book,” remained the responsibility of national authorities. While the Basel procedures is itself a consideration as to whether the Federal Reserve will object to a bank’s capital distribution plans.18 The advantage of stress-tests in discouraging banks from producing low loss estimates is offset by the disadvantage that it encourages banks to use models that produce results similar to those of the Federal Reserve’s model, according to Schuermann (2013). As a result he argues, “if everybody uses the same scenario (for interest rate risk) and works hard to get the same numbers (and they are trying), then we have a very narrowly specialized risk machine that is inflexible and unresponsive to unexpected shocks.” He further argues that deviating from standard industry practice is becoming discouraged, resulting in less innovation even in cases where risk managers see opportunities for more accurate modeling. 4. Interest rate risk mis-measurement Stress-tests could, in principle, provide better interest rate risk measurement because Basel III only incorporates the effect of interest rates on market values for those assets held for trading (the trading book). Currently, there is no interest rate risk measure for assets likely to be held to maturity (the banking book). However, as currently conducted, do not use the wealth of detailed information about assets in the banking book used in the loss analysis. Rather, the Board of Governors of the Federal Reserve (2013, p. 46) reports that its stress-test model for PPNR uses a series of “autoregressive models that relate the components of a BHC’s revenues and non-credit-related expenses to BHC characteristics, and to macroeconomic variables.” It further reports that models are estimated for 17 different components, including 5 components of interest income and 3 of interest expense. The reliance on highly aggregated data and regressions based on historical data would seem to afford considerable opportunity for banks to increase their actual exposure to interest rate changes without showing appreciably higher exposure in stress-tests.

Moreover, even if the stress-tests did model individual items in the banking book, they could at best partially capture the risk of those positions because the stress-tests focus on book values 18 Bloom (2008) shows that the addition of a leverage ratio may also increase the effectiveness of the advanced approach of Basel II in two ways: (a) the leverage ratio reduced the extent to which a bank may reduce its capital requirements by understating risk and (b) a higher leverage ratio makes supervisory threats to fine banks that are found to underestimate risk more credible. Not only can properly designed stress-tests help accomplish the first goal in a more risk sensitive way than the leverage ratio but the enforcement mechanism used in the stress-tests (changes to banks capital distribution plans in good times and requirements to issue new capital in bad times) is credible in a wider variety of circumstances.

19 Becker (2013) reports that efforts are under way at the Basel Committee on Banking Supervision but that these efforts are at a relatively early stage.

20 The guidance and the FAQ were issued as “Interagency guidance” jointly by the Board of Governors of the Federal Reserve System (FRB), the Federal Deposit Insurance Corporation (FDIC), the National Credit Union Administration (NCUA), the Office of the Comptroller of the Currency (OCC) and the State Liaison Committee as Interagency. The 2010 guidance is available at http://www.federalreserve.gov/newsevents/press/bcreg/bcreg20100107.pdf and the FAQ is available at http://www.ffiec.gov/PDF/01-12RR_FAQs.pdf [accessed 21 September 2013].
rather than market values. For example, a shift up in the term structure of interest rates will decrease the value of a 30-year, fixed-rate mortgage loan both because of higher discount rate applied to all of the loan repayments and slower prepayments at a time of high reinvestment rates. However, the current supervisory stress-tests only capture lost income over the next three years, essentially the losses that may arise because higher rates increase the cost of funding but do not increase the rate paid on a fixed-rate mortgage.

Another weakness of stress-tests is that they only measure interest rate risk for the tested scenarios. Banks may have considerable interest rate risk, but that risk may be concentrated in scenarios that are not examined. For example, banks may seek to compensate to boost their spread income in a low interest rate environment by investing in longer maturity assets while holding the maturity of their funding constant. However, if supervisors’ primary concern is credit risk, the stress-tests may focus on scenarios in low or negative economic growth scenarios, which may capture the banks’ exposure to credit risk but which assume continuing low rates and hence miss the banks’ interest rate risk.

Moreover, even if the supervisors try to stress banks’ exposure to interest rate risk, the tests may fail to capture the risk. One way this could happen would be if the supervisors try to stress banks’ exposure to rates by including a scenario with sharply rising rates but some banks make a contrarian bet. That is, some banks position their portfolios to gain from rate increases at the cost of being exposed to rate declines. Another alternative, facilitated by the existence of interest rate options (including embedded options), is that a bank is hedged against large interest rate moves but exposed to losses from smaller moves. For example, if supervisors develop a reputation for stressing banks using a 200 basis point change, a bank might be able to acquire out of the money options that provide an effective PPNR hedge against 200 basis point moves but provide no PPNR protection against 190 basis point moves.

5. Capital mis-measurement

The measures of capital used in Basel capital accords (Basel I, II and III) are taken directly from banks’ financial statements, with only mechanistic adjustments reflecting the impact of various items on a bank’s ability to absorb losses. Unfortunately, financial accounting principles balance a variety of objectives and give bank management responsibility for a large number of judgmental decisions. As result, the supervisory capital ratios (all of which are based on financial accounting measures of capital), lacked credibility during the financial crisis of 2008 (Wall (2013b)).

In contrast, stress-tests are designed to measure changes in capital given a particular scenario. A sufficiently adverse scenario maintained over a long enough period is likely to force a bank to (at least partially) recognize credit losses embedded in their portfolio. The SCAP scenario was adverse and had a two-year horizon, partly to force recognition of the embedded credit losses. As a result, one of the accomplishments of SCAP was to help restore confidence in banks by providing better estimates of the unrecognized losses embedded in their portfolios.

A weakness of stress-tests in measuring capital is that their effectiveness depends in large part upon the stress scenario being tested. A sufficiently favorable scenario may not require a bank to recognize most of the embedded losses in its portfolio. Wall (2013b) emphasizes that U.S. supervisors could choose a stressful scenario being confident that the results would not adversely impact market participants’ confidence in banks’ existing values (the post-Lehman runs suggested that participants already had little confidence in reported equity) and the U.S. supervisors had a mechanism for addressing any capital inadequacies that could not be made good by private issues of new capital (Capital Purchase Program of the Troubled Assets Relief Program, TARP).

Another set of changes that may partially reduce the relative value of stress-tests in measuring capital would be the proposed changes in accounting standards. Wall (2013a) discusses a proposal by the Financial Accounting Standards Board (FASB) that would require banks to recognize expected losses rather

---

21 Admittedly, this is a scenario more easily imagined if short-term rates were several percentage points higher than their current level.

22 See Jarrow (2013) for a discussion of the incentive created by stress-testing to concentrate risk in the extreme tail of the loss distribution.

23 For example, certain intangible assets are deducted from capital reflecting likely declines in the values of these assets in a distress situation.
than incurred losses, which result in earlier loss recognition and eliminate an important source of managerial discretion to delay loss recognition.

6. Countercyclical capital buffers

Capital adequacy rules that require a constant minimum throughout the business cycle are unavoidably procyclical.\textsuperscript{24} Estimates of riskiness for RWA purposes are likely to be at cyclical lows during boom times and additional capital is likely to be readily available, especially additional retained earnings.\textsuperscript{25} As a result, banks are best able to expand their lending and support economic growth when times are already good. Conversely, risk estimates are likely to be higher and losses may be depleting capital during stress periods, reducing bank's ability to lend and possibly leading them to refuse to roll over or make new loans to good customers.

The problem of procyclical capital regulation during the crisis sparked efforts to vary capital requirements in a countercyclical manner to dampen loan growth in boom times and facilitate bank lending during bad economic times. One result was that Basel III was modified to include a countercyclical capital buffer. The size of this buffer is to be determined by national supervisory authorities within a range of 0% to 2.5%.

The questions that national authorities must answer are when to invoke this buffer and how large a buffer should be required of banks. The problem with non-discretionary triggers is that the state of the art for measuring the buildup of risks to financial stability is not very advanced.\textsuperscript{26} The alternative is to give the supervisors discretion in setting the buffer. Supervisory determined buffers are unlikely to be consistently set at optimal levels, but without any external pressure the buffers would likely be varied in ways that are superior to the current fixed buffers. However, a discretionary decision to change the buffer will surely be a high-profile decision that will attract public attention. Supervisors considering a change in the buffer will face considerable external pressure to not increase the buffer in a timely manner (or to prematurely decrease the buffer) from banks, and sectors of the economy benefiting from economic booms.

Stress-tests provide another, not mutually exclusive, mechanism for enforcing countercyclical capital buffers. As Acharya (2012) notes, a way of accomplishing this would be to use an adverse scenario that does not moderate during good times. For example, the assumption of a 10% unemployment rate may imply only a small increase in unemployment during a recession but a rather larger increase in unemployment during a boom. As a result, banks would automatically need to hold a larger capital buffer during good times than bad. Moreover, the use of such a constant scenario over the business cycle would likely put the supervisors in a better position to defend countercyclical capital requirements than a purely discretionary decision under Basel III.

7. Conclusion

Basel III and stress-testing represent significant increases in the sophistication and cost of measuring banks' capital adequacy. This paper discusses a variety of ways in which stress-testing could be used to mitigate problems with Basel III. These problems, including cases where Basel III underestimated credit risk, but not explicitly include interest rate risk, relies on potentially inflated estimates of bank capital and on the use of politically vulnerable discretion for the implementation of its countercyclical capital buffer.

The weaknesses in Basel III can be mitigated by stress-testing because of differences in the way the two measures are structured and implemented. Stress-testing can mitigate the incentives created by Basel III credit risk underestimation for three reasons: (a) stress-test errors are unlikely to be perfectly correlated with Basel III errors, (b) stress-tests are less reliant on models run by banks and (c) stress-test scenarios can be designed to address weaknesses in both the Basel III risk weightings and supervisors' perception of estimation error in the stress-test models. Stress-testing could mitigate the failure of Basel III to include an explicit interest rate risk component by including an interest rate stress scenario in the analysis and using more granular data on individual bank's current exposure. Stress-testing can mitigate Basel III's use of possibly overvalued book capital by forcing banks to estimate losses over a multiyear period

\textsuperscript{24} The actual effect in practice, however, is likely to be somewhat less procyclical as regulatory capital adequacy requirements give banks an incentive to build up capital buffers in good times, which can then be run down during periods of stress.

\textsuperscript{25} See Repullo and Suarez (2012) and Andersen (2011).

\textsuperscript{26} See Edge and Meisenzahl (2011) for a discussion of the unreliability of one proposed measure of buildups, the credit-to-GDP ratio. See also Pritsker (2012) who provides a recent theoretical analysis of the choice of stress-test scenarios.
in scenarios in which economic conditions do not improve enough to make good embedded credit losses. Finally, stress-testing can mitigate Basel III's reliance on the exercise of discretion for countercyclical capital buffers by the use of scenarios that do not become significantly more adverse during downturns.

Given that stress-tests can add value, the question becomes will they add value. The answer depends, to a very considerable degree, upon the stress-test scenarios implemented by banks and supervisors. Identifying the buildup of risky credit exposures before they turn into large losses requires that the supervisors are willing to test seemingly very adverse scenarios at a time when banks are still doing relatively well. Similar to identifying economic losses in a stressful situation, supervisors will require the use of a scenario that could potentially reveal losses that would leave some banks reporting capital well below regulatory minimums over parts of the stress horizon. Although such risk identification ex-ante and loss recognition ex-post are possible using stress-tests, supervisors have not always been particularly aggressive in using their other powers to identify risks or require loss recognition. Whether this supervisory reluctance will change merely because the supervisors are now employing stress-tests is not obvious.27

References


Becker, L., 2013, “Basel Committee has work cut out on interest rate risk charge,” Risk, 17 July

Bernanke, B. S., 2013, “Stress testing banks: what have we learned?” Speech given the Federal Reserve Bank of Atlanta’s Financial Markets Conference, 8 April
Peristian, S., D. P. Morgan, and V. Savino, 2010, “The information value of the stress test and bank opacity,” Federal Reserve Bank of New York Staff Reports No. 460
United States Council of Economic Advisers, 2012, “2012 economic report of the President,” Chapter 4
Wall, L. D., 2013a, “FASB proposes (too?) early loan loss recognition,” Notes from the Vault, Federal Reserve Bank of Atlanta, August
Wall, L. D., 2013b, “The adoption of stress testing: why the Basel Capital measures were not enough,” unpublished working paper

27 Indeed, the 2010 and 2011 European stress-tests already provide an example where stress-tests were designed and implemented in a way that gave a misleadingly favorable impression of many bank’s condition, see Ahmed et al. (2011) and Wall (2013b).
Externalities and macroprudential policy

Gianni De Nicolò
Senior Economist, Research Department, International Monetary Fund, and Fellow, CESifo

Giovanni Favara
Economist, Federal Reserve Board

Lev Ratnovski
Economist, Research Department, International Monetary Fund

Abstract
As for any form of government intervention, macroprudential policy should be justified by market failures. This paper discusses three key externalities across financial institutions and from financial institutions to the real economy that rationalize the need for macroprudential policy: externalities related to strategic complementarities, fire sales and interconnectedness. We link each externality to recently proposed macroprudential policy tools, and argue that although various tools can correct the same externality, these tools are best seen as complements rather than substitutes.

1 We would like to thank Olivier Blanchard, Stijn Claessens, Giovanni Dell’Ariccia, Luc Laeven and Anton Korinek for useful comments and suggestions. All errors are our own. The views expressed in this paper are those of the authors and are in no way representative of the views of the IMF or the Federal Reserve System, their management, or any affiliated organizations. All remaining errors are the sole responsibility of the authors.
1. Introduction

The 2007-08 financial crisis and the ensuing “Great Recession” have led to a profound re-examination of macroeconomic policy, and financial regulation and oversight. At the heart of this re-examination is the realization that financial stability has a critical bearing on macroeconomic outcomes [Blanchard et al. (2010)].

The renewed focus on financial stability has sparked important financial regulation reforms. The new Basel III accord envisions more stringent capital regulation, as well as the introduction of new tools, such as liquidity requirements. Other reforms, affecting banks’ activities, the shadow banking system, and changes in the institutional infrastructure, are under way in various jurisdictions.

Beyond traditional microprudential regulation, the crisis has led to a new focus on “macroprudential” policy, which aims to address systemic risk, i.e., “the risk of developments that threaten the stability of the financial system as a whole and consequently the broader economy” [Bernanke (2009)]. The proposed tools range from adaptations of standard microprudential measures, such as capital surcharges that are countercyclical or systemic risk-based, to taxes and restrictions on intermediaries’ assets and liabilities [FSB (2011a, b), IMF (2011a), and Shin (2011)]. Many countries have announced the adoption of macroprudential policies, and some have already implemented versions of such policies with mixed evidence as to their effectiveness [see Crowe et al. (2011), IMF (2011b), Lim et al. (2011)].

The focus on macroprudential policy is motivated by the fact that microprudential regulation is necessary but not sufficient to deal with systemic risk. Microprudential regulation (such as Basel I and II capital accords) tends to view financial institutions in isolation and aims mainly to ensure that each is individually solvent. Yet solvency of individual institutions is not a sufficient condition for the stability of a system as a whole, for two main reasons. First, the focus on individual institutions neglects risks that are of systemic rather than individual nature, such as correlation risk [Acharya (2009)]. Second, certain aspects of microprudential regulation, while aimed at protecting individual institutions, may at times destabilize the system as a whole [Hanson et al. (2011)]. For example, microprudential capital requirements that become binding following a shock can turn individual deleveraging into a system-wide credit crunch. Given these limits, the purpose of macroprudential regulation is to focus on the financial system as a whole, with an ultimate objective of limiting systemic risk.

This paper aims to advance the debate on macroprudential policy by focusing on its economic rationale, i.e., the correction of market failures that give rise to systemic risk. To do so, we identify three types of externalities: (a) externalities related to strategic complementarities, which lead banks (and other financial institutions) to take excessive or correlated risks and build up vulnerabilities during the upswing of a financial cycle; (b) externalities related to fire sales, arising from a generalized sell-off of financial assets that causes a decline in asset prices and impairs the balance sheet of intermediaries, amplifying the contractionary phase of the financial cycle; (c) externalities related to interconnectedness, caused by the propagation of shocks from systemic institutions or through financial networks.

Correcting these externalities can be seen as intermediate targets for macroprudential policy, since policies that control externalities mitigate market failures that create systemic risk. This approach may help policymakers select the most effective instruments for macroprudential policy. By focusing on the economic rationale, our approach complements the traditional, descriptive classification of macroprudential policies based on the distinction between time-series and cross-sectional dimensions of systemic risk [Borio (2009), Bank of England (2011)].

While it would be desirable to derive from first principles the optimal response to the identified externalities, it is difficult to do so. The narrower approach of this paper is to discuss how the main proposed macroprudential policy tools – variations of capital requirements, liquidity requirements, restrictions on activities, and taxes – may address the identified externalities.

We find that the three externalities identified in this paper may be corrected by various tools, but each tool has benefits and limitations. As a result, various tools addressing the same externality are best seen as complements. For example,

---

2 The term “macroprudential” was put into use by Crockett (2000) and Borio (2003). Galati and Moessner (2011) provide a review of the recent literature on macroprudential policy.

3 See Goodhart et al. (2012) for a recent attempt in this direction.
both capital requirements and limits on bank asset allocation can correct the externalities associated with strategic complementarity. However, capital requirements may become less effective in booms (when capital ratios increase due to buoyant asset prices), so more crude but direct quantity restrictions, such as debt-to-income (DTI) or loan-to-value (LTV) ratios, can supplement them. Similarly, capital and stable funding measures can be complements in addressing the risk of fire sales, since they focus on vulnerabilities on different sides of the balance sheet. And, capital surcharges can weaken the incentives of banks to become systemic and ensure that systemic institutions have a larger buffer to use in case of distress, while restrictions on the composition of bank assets (e.g., as envisioned by the Volcker rule) can limit their most risky exposures.

Despite these complementarities, one conclusion that emerges from this paper is that capital surcharges are likely to play an important role in any macroprudential framework. The reason is that relative to other policy tools, time-varying and systemic risk-based capital surcharges address any of the three identified externalities that lead to systemic risk. Moreover, they are closely linked to microprudential regulation and are already part of the Basel III accord.

In this paper, we also suggest that a key challenge for the design of macroprudential policy is to acquire more evidence on the effectiveness of alternative policy tools. While policymakers have so far relied mostly on “rule of thumb” considerations, it is necessary to gain evidence on the effectiveness of interventions in order to make the macroprudential policy design more precise.

Since the analysis in this paper is limited to ex-ante regulation of financial intermediaries, we abstract from other important challenges related to macroprudential policy. For example, we do not cover issues related to the supervision of financial intermediaries. We also take as given the imperfect mechanisms to resolve distressed financial institutions [Claessens et al. (2011)] and abstract from systemic risk implications of capital flows and monetary and fiscal policies [Ostry et al. (2011), De Nicolò et al. (2010), Moolj (2011)].

The structure of the paper is as follows. Section 2 reviews the rationale for microprudential and macroprudential regulation. Section 3 discusses the three externalities that lead to systemic risk. Section 4 links these externalities to macroprudential policy tools and discusses policy design challenges. Section 5 concludes.

2. Why is traditional microprudential policy not enough?
The prevailing economic rationale for microprudential regulation rests largely on a standard moral hazard argument. Shareholders of leveraged firms have incentives to engage in risky activities, since they reap the benefit of the upside, but in the event of failure creditors bear the cost. This classical risk-shifting problem is worse for banks because of public safety nets. Deposit insurance, intended to prevent bank runs and panics, worsens the incentives of depositors to monitor bank risk strategy. The expectation of government bailout, which may be necessary to limit the ex-post cost of financial distress, similarly reduces the monitoring incentives of uninsured creditors.

For this reason, and given the crucial role that debt plays in the capital structure of banks, capital requirements have been a central feature of microprudential regulation. Equity increases the shareholders’ “skin in the game,” forcing them to bear a greater share of the cost of the bank’s risk-taking choices.⁴

The recent financial crisis has shown, however, that solving the moral hazard problem at each individual institution is not a sufficient condition to ensure the stability of the financial system as a whole. One lesson of the crisis is that interconnectedness among banks and other financial institutions can generate externalities with adverse effects on the real economy. A common source of such externality is contagion, the possibility of one distressed bank affecting the stability of others. But the links can be more subtle and involve, for example, pecuniary externalities through asset prices (asset liquidations that result in price declines impairing the balance sheets of other institutions) or strategic interactions among banks that lead them to take correlated risks ex ante. Indeed, the crisis has brought to the fore the idea that the risk in a

---

⁴ A complementary view of prudential regulation and capital requirements is that because small and inexperienced depositors are unable to monitor banks’ manager behavior, regulation is needed to guarantee that banks act in the interest of depositors (Dewatripont and Tirole (1994)).
The financial system is not simply the aggregation of individual risks, but is mostly endogenous risk, resulting from the collective behavior of financial institutions.\(^5\)

Another lesson of the crisis is that microprudential regulation is not always suited to address risks that are systemic in nature. In fact, ensuring the stability of each institution individually can at times destabilize the system as a whole. For example, capital ratios set independently of the business cycle may be a source of systemic risk. Since it is costly to raise new capital in downturns, banks hit by a negative shock may prefer to delever. In aggregate, this collective behavior may cause a credit crunch and a generalized drop in asset prices, exacerbating the initial negative shock.

In the presence of such externalities, microprudential regulation needs to be supplemented with tools that aim to safeguard the financial system as a whole. It is common to refer to such tools as macroprudential policies, although there is no general agreement on their definition and objectives. The standard view is that macroprudential policies address the time and cross-sectional dimensions of systemic risk in the financial system. This classification implies that the objective of macroprudential policies is to smooth financial and credit cycles, in order to prevent systemic crises and provide cushion against their adverse effects.

In this paper, we take the view based on the “first principles” that regulation needs to be justified by market failures. This approach clarifies that macroprudential policies are justified by the need to correct market failures, and not simply because the financial system is “fragile.” This provides a justification for regulation, and a framework to analyze the economics behind recent policy proposals.

### 3. What are the externalities — market failures that justify macroprudential policy?

We start our discussion by identifying market failures that may require a response in the form of macroprudential policy. The recent financial crisis and the debate it spurred in the literature highlight three main sources of such market failures: externality related to strategic complementarities, fire sales and interconnectedness.\(^6\) Externalities operate between financial institutions, and therefore cannot be fully addressed by microprudential policy that views institutions in isolation. In principle, there may be other market failures that require a macroprudential policy response; the scope of our discussion is confined to these three externalities, reflecting the current state of knowledge.\(^7\)

3.1 Externality related to strategic complementarities

Historical experience suggests that financial intermediaries tend to assume exposure to common credit and liquidity risk in an upswing of a business cycle, which amplifies credit and liquidity cycles and generates asset price volatility.

Banks and other financial intermediaries choose to correlate their risks because of strategic complementarities, meaning that the payoff from a certain strategy increases with the number of other agents undertaking the same strategy. Some complementarities are driven by the simple market interactions of rational agents, whereas others stem from the optimal ex-ante response of agents to ex-post government intervention in the event of a financial crisis. We refer to the effects of these strategic complementarities as externalities.

One source of strategic complementarity is related to increased competition in boom times, which can affect economy-wide credit standards. In the presence of imperfect information, banks’ incentives to assess borrower risk depend on their strategic interaction. In booms, banks have fewer incentives to screen potential borrowers due to lower rents prompted by fiercer competition. As a result, they reduce screening intensity and increase lending. This causes a worsening of the pool of borrowers, which is reversed through lower credit origination and milder competitive pressures once the contractionary phase sets in [Ruckes (2004), Dell’Ariccia and Marquez (2004), Gorton and He (2008)].

A second strategic complementarity is caused by reputational

---

5 Brunnermeier et al. (2009) frame the distinction between microprudential and macroprudential regulations as a “fallacy of composition”; safeguarding the components of the financial system is not a sufficient condition to safeguard the system.

6Externalities have welfare effects in the presence of asymmetric information or market incompleteness [Greenwald and Stiglitz (1986)]. The papers reviewed in this section identify externalities in the financial system in the presence of such market frictions.

concerns and the incentives structure facing bank managers. When bank managers care about the market perception of their abilities, their credit policies are influenced by those of other banks [Rajan (1994)]. Benchmarking creates externalities across banks because a bank reporting poor performance due to losses will be evaluated more leniently by the market if many other banks suffer loan losses at the same time. Banks, therefore, have incentives to hide losses or maintain risky lending until the buildup of bad loans forces them to “coordinate” a strategy of loss recognition and credit contraction.

The prospect of a government bailout in the event of financial distress can also lead banks to engage in correlated asset choice ex-ante. Anticipating that simultaneous bank failures trigger a bailout (to prevent a financial meltdown), banks may find it optimal to correlate risk to maximize the probability that any failure is a joint failure [Farhi and Tirole (2011), Acharya and Yorulmazer (2007)]. As firms mimic each others’ strategy, vulnerabilities in the financial system increase.

Complementarities can affect banks’ asset choices, or manifest themselves in the form of excess maturity or exchange rate mismatches [Ratnovski (2009), Allen and Carletti (2011)]. The upshot of these mechanisms is that the quality of banks’ and other investors’ portfolios worsens in a boom, and banks become exposed to the same types of risk, creating vulnerabilities that lead to or deepen the downside of the financial cycle.

3.2 Externalities related to fire sales
Fire sales typically arise in downturns, amplifying financial distress through a pecuniary externality. A fire sale occurs when a financial institution is forced to liquidate an asset at a time when potential buyers are also troubled. Given limited potential buyers, the asset is sold at price below its fundamental value, causing losses to the seller [Shleifer and Vishny (1992), Allen and Gale (1994)]. Not only does this asset fetch a lower price, but also similar assets held by other banks may decline in value. This reduces the capital ratios of these banks and their ability to post assets as collateral, forcing them to liquidate underpriced assets. The new round of selling triggers further losses, new selling, etc. A fire sale can bring multiple banks to financial distress, and through this trigger a credit crunch with adverse real consequences.8

The forced sale condition is an obvious possibility for banks because one of their main functions is to issue liquid liabilities to fund investments in illiquid assets. This maturity transformation exposes banks to the risk of having to liquidate investments prematurely in case of a sudden withdrawal of funding. Although government guarantees, such as deposit insurance and liquidity facilities (discount window), reduce the likelihood of fire sales, their effectiveness is limited when banks also rely on wholesale funding or when other important players in the intermediation process, such as broker-dealers and “shadow banks,” do not (formally) benefit from such government guarantees and liquidity support.9

Although the externalities associated with fire sales manifest themselves in a downturn, the imbalances that sow the risk of fire sales are often built up in booms. The reason is that atomistic agents take prices as given, but in aggregate the equilibrium price depends on their joint behavior. As a result, private agents may overborrow, leading to excessive leverage and inflated asset prices, because they do not internalize the costs that a generalized fire sale may have on the ex-post borrowing capacity of other agents [Caballero and Krishnamurthy (2003), Lorenzoni (2008), Korinek (2011) and Stein (2012)].

3.3 Externalities related to interconnectedness
Banks operate in an interconnected system. As a result, distress or failure of a bank can affect other institutions. Spillovers can arise because of asset price movements [as discussed in the previous section, see also Nier et al. (2007)], bilateral interbank market exposures [Allen and Gale (2000), Diamond and Rajan (2005)], or feedback from the real economy [Bebchuk and Goldstein (2011)].

Banks can reduce but not entirely eliminate contagion risk for two reasons. First, the shape of interconnectedness in

---

8 There are two additional channels through which fire sales can undermine investment. First, if the price dislocation is extreme, with asset prices dropping extensively, banks may find it convenient to use spare financial capacity to buy the extreme underpriced assets instead of lending to firms [Shleifer and Vishny (2010)]. Second, the prospect of an ex-post fire sale may lead banks to take precautionary measures and hold ex-ante liquid assets instead of lending [Diamond and Rajan (2012)].

9 Institutions that make up the shadow banking system are money market mutual funds, repo-financed dealer firms, and securities lenders. Shadow banks commonly obtain short-term funding in wholesale markets and invest these funds in longer-term financial assets, performing maturity transformation that is similar to that of banks [see, e.g., Gorton and Metrick (2011), Pozsar et al. (2010)].
the financial system is beyond the individual bank's control.\footnote{For example, a bank has no control over other interbank exposures of its counterparties [Acemoglu et al. (2012)].} Second, interconnectedness may arise naturally as a result of mutual hedging and diversification motives [Wagner (2011), Allen et al. (2012)]. Banks may become overly interconnected when they do not internalize the implications of their own interconnectedness on systemic risk, or behaviorally “neglect” the possibility of rare but large shocks [Acemoglu et al. (2012), Gennaioli et al. (2011)]. The growing financial networks literature [Allen and Gale (2000), Allen and Babus (2009), Gai et al. (2011)] suggests that high interconnectedness mitigates the impact of small shocks by spreading them, but amplifies large shocks since they can reach more counterparties. Externalties stemming from interconnectedness are particularly strong for systemically important financial institutions (SIFIs). Unlike smaller institutions, distressed SIFIs cannot be easily wound down, since they are complex, operate internationally and play a role as backbones of the financial infrastructure. Historically, most interventions in SIFIs were de facto bailouts, which have protected their shareholders and creditors from the full scale of distress-related losses. The anticipation of bailouts has the adverse effect of increasing risk-taking incentives. It also reduces market discipline and effectively subsidizes SIFIs, especially the riskiest ones [O'Hara and Shaw (1990), Flannery (2009), Ueda and Weder di Mauro (2011)]. In addition, the near-certain government support to distressed SIFIs introduces a race among financial institutions to become systemically important, as this implies a lower cost of funding in normal times and a better protection against losses when in distress.

While historically “systemic importance” has been associated with institutions’ size, recent events suggest a more complex picture. The interconnectedness of a SIFI is also determined by its interbank market linkages, and its effects are amplified by high leverage [Drehmann and Tarashev (2011a)]. Interconnectedness may also be present in non-banks (e.g., hedge funds or money market mutual funds), or institutions that support market infrastructure, such as central clearing counterparties.

4. How to link theoretical market failures to practical policy tools?

The ultimate goal of macroprudential policy is to reduce “systemic risk”, i.e., risks that threaten the stability of the financial system as a whole, and can adversely impact the real economy. This broad definition of systemic risk is not, however, well suited for the formulation of macroprudential policy. The reason is that it does not highlight the sources of systemic risk – market failures.

The approach taken in this paper allows us to treat the correction of externalities as the intermediate target for macroprudential policy. Policies that control externalities correct market failures that create systemic risk. This approach gives more structure to the definition of systemic risk, and introduces economic rationale into the discussion of macroprudential policy.

It would be ideal to derive from first principles optimal policies that correct the externalities. Unfortunately, a comprehensive theoretical treatment is difficult. At least since Weitzman (1974), it has become clear that the optimal policy rests on assumptions regarding the set of instruments and information available to the regulator [Kaplow and Shavell (2002)]. Heterogeneity of agents and their capacity to engage in moral hazard also play a role [Perotti and Suarez (2011)]. For the financial sector, all four dimensions are complex. Information is often asymmetric, and its production requires effort by the regulator and market participants [Flannery (1998)]. Instruments include not only taxes and (various) quantity restrictions, but also hybrids: capital and liquidity requirements combine the properties of price (in liquid markets) and quantity tools (in illiquid markets) and provide buffers to offset losses in financial distress. Agents have different comparative advantages, including in their capacity to assess, monitor, diversify and bear risks [Boot (2000)]. And finally, there is significant scope for regulatory arbitrage, ability to conceal risks, and moral hazard more generally. Incorporating all these elements into a single framework in a meaningful way is likely impossible. Analytical treatment has to be selective, looking at specific contexts and trade-offs.

For this reason, we have decided to adopt a more “practical” approach. We consider the existing set of macroprudential policy proposals [see Bank of England (2011) and Lim et al. (2011) for overviews] and assess their advantages and disadvantages.
in correcting the three externalities discussed in Section 3. Specifically, we map the externalities discussed in the previous section with the following key macroprudential policy proposals: enhanced capital requirements, liquidity regulation, restrictions on bank activities and asset allocation, and taxation.\textsuperscript{11} Table 1 provides a simplified representation of this mapping.

Most of the instruments considered in this table (e.g., capital or liquidity requirements, and restriction on assets and liabilities) are akin to tools already used in traditional microprudential regulation. What gives these instruments a macroprudential flavor is that they are not imposed to resolve agency conflicts within a bank, but rather to correct externalities that arise in between banks. Consequently, these measures do not depend on bank characteristics taken in isolation (as for microprudential regulation), but are contingent on the aggregate behavior of all banks (e.g., the lending cycle) or the position of a bank within the financial system (e.g., systemic importance).\textsuperscript{12}

In what follows, we explain how the instruments outlined in Table 1 can correct the three externalities. Although these instruments may be seen as substitutes, they often entail different advantages and limitations, and become complements in addressing the same externality.

4.1 Correcting externalities related to strategic complementarities
Of the policies considered in Table 1, two address the externalities related to strategic complementarities: capital requirements and restrictions on bank asset allocation. Both tools limit banks’ expected gains from choosing correlated lending strategies: capital requirements induce banks to internalize more of the cost of engaging in risky lending; restrictions on asset allocation prevent banks from taking large exposures. These tools can thus limit asset growth in the upturn of the credit cycle so as to reduce the cost associated with adjustments in downturns.

Time-varying capital requirements, in the form of a capital surcharge linked to aggregate credit growth, are part of the new Basel III accord [Basel Committee (2011)]. Basel III also allows for adjusting risk weights in order to control exposures to specific assets, such as real estate loans. The time-varying feature of this surcharge is new relative to standard microprudential regulation, and is meant to mitigate risk-taking during credit expansions.\textsuperscript{13}

Several countries have also imposed restrictions on bank assets through caps to loan-to-value (LTV) or debt-to-income (DTI) ratios. These restrictions affect directly the asset side of a bank’s balance sheet and are meant to limit the fall in banks’ lending standards in booms. While similar quantity restrictions historically were widely used as part of macro-stabilization policies to control credit flows, limits to LTV and DTI ratios have a macroprudential purpose because they are designed to bind in expansions.

Although capital and restrictions on assets can independently mitigate externalities associated with strategic complementarity,

\begin{table}[h]
\centering
\begin{tabular}{|l|c|c|c|}
\hline
\textbf{Externality} & \textbf{Capital requirements} & \textbf{Liquidity requirements} & \textbf{Restrictions on activities, assets or liabilities} & \textbf{Taxation} \\
\hline
\textbf{Strategic complementarities} & X & & & \\
\hline
\textbf{Fire sales} & X & & X & \\
\hline
\textbf{Contagion} & & X & X & \\
\hline
\end{tabular}
\caption{Externalities and macroprudential policies}
\end{table}

\textsuperscript{11} Dynamic provisioning and leverage ratios – two commonly discussed macroprudential tools – may be considered as capital-based instruments. In our analysis, we omit the discussion of macroprudential policy tools that target market functioning and infrastructure, e.g., through-the-cycle margining, central clearing counterparties and disclosure of risk.

\textsuperscript{12} The design of macroprudential policy may conflict with microprudential objectives. For example, macroprudential capital requirements are high in booms (to limit the accumulation of imbalances) and low in recessions (to avoid fire sales and deleveraging). In contrast, optimal microprudential capital requirements are low in booms (when the volume of defaults is low) and high in recessions (when banks need more equity to absorb an increase in defaults). In downturns, pursuing solely a macroprudential objective may lead to instability of individual banks, while a focus only on microprudential objective may lead to a deleveraging spiral at the system level.

\textsuperscript{13} The economic effects of procyclical capital can also be replicated using dynamic provisioning, cf. Saurina (2009).
they have different drawbacks. Capital requirements are useful for incentives and as a buffer, but can become less effective in booms, when capital ratios increase due to high profitability and buoyant asset prices [Shin (2011)]. For this reason, quantity restrictions, such as DTI and LTV ratios, can complement capital requirements by imposing direct constraints on asset allocation. They can also target specific risky borrowers and asset classes that raise macroprudential concerns.14

4.2 Correcting externals related to fi re sales
The externals associated with fire sales arise because banks fail to internalize the consequences of not taking precautionary measures in normal times, and thus need to adjust by shedding assets ex-post in the event of a negative aggregate shock. Time-varying capital requirements that are higher in upturns alleviate this externality, by mitigating the incentives for risk-taking in booms. They also provide a buffer that offsets losses and reduces the risk of selling assets at fire sale prices in downturns.15

But, fire sales can also be triggered by disruptions to bank funding markets. As witnessed during the recent financial crisis, bad news about asset values can trigger run-like reactions of creditors that stop rolling over short-term funding, forcing banks to shrink balance sheets. Accordingly, Basel III proposals to impose liquidity and stable funding requirements can be thought of as tools to limit the risk of fire sales stemming from bank reliance on short-term debt. A related way to limit the use of short-term debt is to rely on taxes calibrated to the cost difference between long- and short-term funding [Huang and Ratnovski (2011), IMF (2010a), Perotti and Suarez (2011)].

Capital and funding measures can be seen as complements, since they affect fire sale risks on different sides of the balance sheet. Less obvious is the optimal choice between liquidity requirements and the taxation of unstable funding. Basel III gives preference to liquidity requirements. But, there are sound arguments for taxing bank funding risk. Taxes can be easily calibrated to reflect the price difference between stable and unstable funding, and offer more flexibility during stress periods, when banks have to “eat into” minimal liquidity or accept less stable funding.

4.3 Correcting externals related to interconnectedness
The incentives of banks to become systemically important and to take excessive risk can be addressed through effective resolution mechanisms. However, resolution procedures remain imperfect nationally and especially for systematically important cross-border banks [Claessens et al. (2010)]. Accordingly, an ex-ante regulatory response that aims to reduce the probability of SIFIs’ distress is necessary. Three of the regulatory tools reported in Table 1 can achieve this objective: capital surcharges, restrictions on asset composition and taxation.

Capital surcharges linked to a measure of systemic importance [BIS (2011)] aim to reduce incentives to become systemically important, as they increase the cost of funding.16 They also generate additional buffer that a SIFI can use in case of distress. A downside of using capital surcharges as a tool for dealing with the systemic risk created by SIFIs is that they are hard to calibrate because it is difficult to map surcharges to SIFI contributions to systemic risk.17 In addition, by making the list of SIFIs public, the surcharges can weaken creditor discipline, exacerbating the SIFI problem.

Complementary tools are restrictions on the composition of bank assets – as envisioned by the Volcker rule to limit proprietary trading or the recommendations of the U.K. Independent Commission on Banking (so-called Vickers Commission) to severe links between retail and investment banking operations. These tools are intended to limit the implicit subsidy of SIFI funding, and to provide a firewall protecting systemically important bank operations from risky activities. The key challenge in implementing restrictions on activities is that the optimal scale and scope of a bank is difficult to determine. Hence, restrictions on activities target only the most evident sources of risk, and can, at best, complement capital requirements. A “softer” form of

14 Some tax-based measures, such as the reduction of tax shield on corporate debt and mortgages and progressive income taxes, can also reduce banks’ incentives to take excessive leverage and risk concentration [Blanchard (2009), IMF (2010a), Lander and Plantin (2011)]. While useful to reduce such risks, these tax instruments do not correct directly the source of the externality due to strategic complementarity, i.e., banks’ incentives to benchmark or to reduce lending standards (see Section 3.1).

15 For non-bank financial institutions engaged in market-based activities, macroprudential regulations can take the form of procyclical margin requirements [see, e.g., Geanakoplos (2009), Gorton (2009), Gorton and Metrick (2010)].

16 This effect can be offset when a designation of an institution as systemically important lowers the required return on its debt and equity.

17 See IMF (2010b) for a discussion of the calibration of systemic risk-based capital surcharges.
restrictions on activities is “living wills” – a document describing bank dissolution and resolution procedures.

The academic debate has also suggested that systemic risk can be addressed through Pigouvian taxes. Such taxes, based on a measure of systemic risk externalities, can force a SIFI to internalize the systemic risk it creates. While theoretically appealing, the implementation of Pigouvian is complicated because it is difficult to measure with sufficient precision the systemic risk contributions of financial institutions [IMF (2010a), Kocherlakota (2010); see also Section 4.5].

4.4 The central role of capital-based tools
The discussion so far suggests that the instruments highlighted in Table 1 – capital requirements, liquidity requirements, restrictions on bank activities, and taxes – have distinct properties, and tend to be complementary for addressing the same externality. Accordingly, no single instrument is a priori a “silver bullet” in correcting externalities. A combination of instruments may seem, instead, more appropriate.

Nevertheless, it appears that capital requirements are likely to play a key role in any macroprudential framework. As discussed above, tools based on time-varying and systemic risk-based capital surcharges can play a role in addressing any of the three externalities leading to systemic risk. Moreover, capital-based tools are closely linked to microprudential regulation and are part of the Basel III accord.

This means that banks – especially the systemic ones, and more so during booms – may have to maintain significantly higher capital buffers than presently. But what are the costs of higher capital requirements? The prevailing view in the literature is that, in the long run, the impact of higher bank capital on the cost of loans is likely to be small [Kashyap et al. (2010), Admati et al. (2010), Herring (2011), Mehran and Thakor (2011)]. The transition to higher capital may, however, be costly. Such costs may be lower if higher capital requirements are introduced gradually, allowing banks to use retained earnings and to time the market for any new equity issuance.18

4.5 Externalities and systemic risk measurement
Systemic risk is a multifaceted phenomenon. As such, there is a variety of metrics that either signal the gradual buildup of imbalances or flag the concentration of risk within the system (Borio and Drehmann (2009), Bisias et al. (2012)).

Each of these metrics captures some of the contributions of the externalities considered in this paper to systemic risk. For example, deviations from trend (gaps) in the credit-to-GDP ratio, property prices, risk premia or leverage can be used to identify externalities related to strategic complementarities or imbalances that can lead to fire sales in downturns [Drehmann et al. (2011), IMF (2011b), Dell’Ariccia et al. (2012)]. Measures of systemic risk contribution, such as CoVaR [Adrian and Brunnermeier (2010)] and systemic shortfall [Acharya et al. (2010), Drehmann and Tarashev (2011b)] can proxy externalities related to interconnectedness. The degree of interconnectedness can also be captured by stress-tests [Brunnermeier et al. (2010), Duffie (2011)] and forward-looking risk indicators [Capuano (2008), Gray and Jobst (2011)].

Although useful to highlight vulnerabilities, these metrics are not sufficient to formulate a policy response. For example, high mortgage credit growth may be caused by a deterioration of lending standards or signal risk concentration. Accordingly, the optimal tools to address the two underlying problems might be different.

Another major challenge is the calibration of macroprudential policy instruments. There is recent evidence that some tools are useful in reducing systemic vulnerabilities, but too little is known quantitatively.19 For now, policymakers have based macroprudential policy on “rule of thumb” considerations. Yet, to make the policy design more precise, it is necessary to gain more evidence on the effectiveness of macroprudential policy tools. For example, how high should capital surcharges be? What is the optimal level of LTV ratio? Further, fundamental and applied research on the optimal choice and calibration of macroprudential policy tools is required to justify policy intervention and avoid regulatory discretion.

18 Flannery (2009) and French et al. (2010) argue that the tax shield costs of higher equity can also be offset using convertible instruments, such as contingent capital.

19 Crowe et al. (2011), Lin et al. (2011) and Dell’Ariccia et al. (2012) evaluate the effectiveness of some macroprudential policy tools.
5. Conclusions

In this paper, we have argued that the first step in the economic analysis of macroprudential policy is the identification of market failures that contribute to systemic risk. Externalities are an important source of such market failures, and macroprudential policy should be thought of as a policy that attempts to correct these externalities.

Building on the discussion in the academic literature, we have identified three externalities that lead to systemic risk: externalities due to strategic complementarities, fire sales and interconnectedness. Externalities due to strategic complementarities contribute to the accumulation of vulnerabilities during the expansion of the financial cycle. Externalities due to fire sales and interconnectedness exacerbate negative shocks.

While it is desirable to derive from first principles an optimal policy that corrects these externalities, it is difficult to do so. In this paper, we have adopted a “practical” approach of considering how externalities can be corrected by the currently proposed macroprudential policy tools: capital requirements, liquidity requirements, restrictions on bank activities, and taxation. We have argued that some of these policies can complement each other in correcting the same externality, and reached the conclusion that capital requirements are likely to play an important role in any macroprudential framework.

We have also argued that although externalities can be proxied through a variety of risk measurements, the accumulation of evidence on the effectiveness of alternative policy tools remains the most pressing concern for the design of macroprudential policy.

References


Blanchard, O., G. Dell’Ariccia, and P. Mauro, 2010, “Rethinking macroeconomic policy,” IMF Staff Position Note 10/03


Crockett, A., 2000, “Marrying the micro- and macroprudential dimensions of financial stability,” BIS Speeches, 21 September


Dell’Ariccia, G., D. Igan, L. Laeven, and H. Tong, 2012, “Policies for macrofinancial stability: options to deal with credit booms,” IMF Staff Discussion Note


Financial Stability Board (FSB), 2011a, “Macro-prudential policy tools and frameworks: progress report to G20,” October, Basel, Switzerland

Financial Stability Board (FSB), 2011b, “Shadow banking: strengthening oversight and regulation,” Basel, Switzerland, October


Flannery, M. J., 2009, “Stabilizing large financial institutions with contingent capital certificates,” mimeo, University of Florida


International Monetary Fund, 2010a, “A fair and substantial contribution by the financial sector: report to the G20,” IMF, Washington


International Monetary Fund, 2011a, “Macroprudential policy: an organizing framework,” Board Paper, April

International Monetary Fund, 2011b, “Towards operationalizing macroprudential policies: when to act?” Chapter 3 in Global Financial Stability Report, September


Directors’ and officers’ insurance and shareholder protection

M. Martin Boyer
CEFA Professor of Finance and CIRANO Fellow, HEC Montréal, Université de Montréal

Abstract
Corporate directors are liable for the corporation’s actions as well as their own. Strangely, and by far, the most likely plaintiffs in a lawsuit against corporate directors are the shareholders who appointed them in the first place. As a result, directors often require protection so that their personal wealth is not expropriated in the event of a good faith error. There are three ways to protect a director’s wealth: corporate indemnification plans, limited liability provisions and directors’ and officers’ (D&O) insurance policies. Of the three types of protection, D&O insurance is arguably the strangest not because shareholders purchase it to protect directors in case of a lawsuit, but because it also protects shareholders. Using an original database, I test a set of hypotheses that should determine the demand for D&O insurance. My analysis suggests that D&O insurance protects the shareholders’ wealth more than the directors’.

1 I conducted much of this research while visiting the Finance division at the Sauder School of Business of the University of British Columbia. I wish to thank Murray Carlson, Gilles Chemla, John Core, Mark Huson, Kai Li, Hernan Ortiz Molina, Randall Morck and Ralph Winter for their comments on this research, as well as seminar participants at the University of Alberta, the University of British Columbia and the University of Connecticut. And, I cannot dismiss the valuable discussions and the continuing work I have on this topic with Sharon Tennyson and Léa Stern. This research was sponsored by the Social Science and Humanities Research Council of Canada. The continuing support of CIRANO is also gratefully acknowledged.
1. Introduction

1.1 Economic Importance of D&O Insurance

As representatives of the corporation, directors and officers are personally responsible and liable for actions committed in its name. One way for corporate directors to protect their personal wealth is to have the corporation buy insurance on their behalf. This insurance is known as D&O insurance. According to different Towers-Watson surveys, the proportion of U.S. firms that carry D&O insurance hovers around 95% throughout the years. The market penetration of D&O insurance is smaller in Canada and is approximately 80%. One possible reason why D&O insurance has become increasingly popular is that lawsuits against management are becoming more frequent. Recurring surveys of Towers-Watson show that one-fifth of all American public corporations had at least one lawsuit brought against their directors in the previous 10 years. Although carrying D&O insurance is the norm, the sum of all D&O insurance premiums paid by U.S. corporations is only a small proportion of the billions of dollars paid in corporate insurance each year. The role of D&O insurance is to cover managers in case they are sued as representatives of the corporation or for breach of duty toward the firm's stakeholders (see Bradley and Shipani 1989), for more on the topic. Managers are covered for their court expenses as well as for any settlement arising from the lawsuit, subject of course to a maximum amount (the policy limit). As D&O insurance protects managers against liability lawsuits brought unto them as managers of the corporation, this is not too different from other types of insurance. What is surprising, however, is that most lawsuits originate from the firm's stakeholders. Lawsuits usually originate from shareholders (close to 50%), employees (30%) and clients (15%). In a sense, shareholders purchase insurance for managers to protect them against shareholder lawsuits. Allowing shareholder to sue the firm and its managers is important as it makes it possible to deal with classic agency problems, since lawsuits act as a disciplinary force on managers and directors (see Becht et al. 2003, and Bauer et al. 2008).

D&O insurance is only one of the three ways that a corporation can protect directors' wealth while they serve on corporate boards. A corporation could also amend its charter so that the directors' liability is limited. Limited liability provisions (LLPs) all but eliminate the directors' personal financial responsibility toward the firm and its shareholders. LLPs became necessary after the 1984 landmark court decision known as Smith versus Van Gorkom (see Hartmann and Rogers 1991, and Hanks 1988, for more details on the case and its impact). Gutierrez (2000) reports that more than 70% of American corporations adopted limited liability provisions in an effort to attract the best possible corporate directors. Brook and Rao (1994) show that firms in financial distress (i.e., the firms whose directors are more likely to be sued) gain from adopting LLPs. Presumably, LLPs allow managers in financially distressed firms to adopt more frequently hero-or-zero strategies when they know they cannot be sued in case of a bankruptcy. Because security litigation settlements may be quite large, and because directors are personally responsible for these amounts, directors seek protection to prevent the use of their personal assets to compensate plaintiffs in these lawsuits. D&O insurance supposedly provides such protection, so individuals require it before they agree to serve as a corporate director.

A second way that a corporation can protect its directors is through corporate indemnification plans (CIPs). CIPs give directors a certain protection against a third-party lawsuit. CIPs are mandatory in the sense that corporations are legally responsible to indemnify directors for court costs that result from a third-party lawsuit if the directors are found innocent of the allegations made against them. CIPs may also be available to the directors even if they are found guilty if, for example, the prejudice was caused in the best interest of the corporation. If that is the case, a director could be indemnified for court expenses and damages. Given that LLPs are frequently part of the corporate charter and CIPs are mandatory, it is not clear why directors would need D&O insurance on top of it all. LLPs limit the amount a director may be responsible for, and CIPs pay for damages to third parties.

Why do corporations purchase D&O insurance on behalf of their directors instead of letting directors purchase such liability insurance directly on the marketplace, just as medical malpractice insurance? I argue in this paper that D&O insurance is not designed to protect directors as much as it is designed to protect the shareholders' wealth. Put differently, D&O insurance acts as a deep out-of-the-pocket put option for shareholders (see also Romano 1991a, Gutierrez 2003, and Boyer and Tennyson 2012). According to Romano (1991a), the largest benefactors of D&O insurance are not the directors, but the shareholders. Directors are already protected under LLPs and CIPs. As a result,
shareholders purchase D&O insurance on behalf of their directors in case shareholders, who feel they were wronged, decide to sue the directors for improper conduct as their representatives on the board. Similar to Romano (1991a), Gutierrez (2003) also sees D&O insurance as an instrument that benefits shareholders instead of directors. Gutierrez suggests that D&O insurance works as a commitment device for efficient shareholder litigation. This shareholder protection hypothesis could also explain why D&O insurance is so widely used to protect corporate directors in common law countries (such as the U.S., Canada¹ and the U.K.). Indeed, if common law countries are better at protecting shareholder rights [see Shleifer and Vishny (1997), and LaPorta et al. (1998, 2000)], why is D&O insurance so prevalent given that it reduces director accountability? One reason may be that D&O insurance protects shareholders more than managers.

Other explanations exist as to why corporations purchase D&O insurance on behalf of its directors. Core (1997) offers two such explanations: managerial risk aversion and managerial compensation [see also Parry and Parry (1991)]. D&O insurance also works as a monitoring tool as insurers are asked to monitor the directors [see O’Sullivan (1997), Holderness (1990), and Core (2000)]. Although risk aversion, compensation and monitoring may appear to be good reasons to purchase D&O insurance, corporations have access to better instruments to achieve those goals: risk aversion is best dealt with by amending the corporate charter to include limited liability provisions; managerial compensation is better served with option, stock and cash distribution; and monitoring is better done by having large shareholders.

1.2 Findings
My findings are based on an original dataset of publicly traded Canadian corporations that filed their annual reports between 1 January 1993 and 31 December 1998. Canadian data is used because the Ontario Securities Commission that oversees the Toronto Stock Exchange mandates that basic D&O insurance information (policy limit, deductible and premium) be made public in the corporations’ management proxies and information circulars. One of the most interesting results of the paper is that I test the Romano (1991a) and Gutierrez (2003) hypothesis that D&O insurance acts as a deep-pocket payer of last resort for shareholders and debt holders. Controlling for other factors, such as director risk aversion and board monitoring, I find evidence that greater shareholder and debt holder wealth leads to greater demand for D&O insurance, thus supporting the shareholder protection hypothesis.

I test seven hypotheses related to the demand of corporate risk management in general and of D&O insurance in particular. A summary of our empirical results follows:

1. If D&O insurance acts as a deep-pocket payer of last resort for shareholders and debt holders, then D&O insurance should be associated with greater shareholder and debt holder wealth. The evidence strongly supports the shareholder protection hypothesis.
2. If D&O insurance is part of the directors' compensation package, we should expect to see a negative relationship between the amounts of D&O insurance protection provided to board members and other forms of compensation. My evidence strongly supports the compensation package hypothesis.
3. If D&O insurance is used to attract more risk adverse directors, we should expect a positive relationship between the board members’ risk aversion and the level of protection. I find no support for the risk aversion hypothesis.
4. If D&O insurance is used to reduce the cost of financial distress, then a positive relationship should exist between the level of D&O insurance coverage and the riskiness of the firm. I find weak evidence in favor of the financial distress hypothesis.
5. If D&O insurance is used as a monitoring device, then we should observe a negative relationship between the amount of D&O insurance coverage and other forms of corporate monitoring. I find weak evidence in favor of the monitoring hypothesis.
6. If D&O insurance provides a signaling mechanism for managerial quality, firms where managerial signaling is more important should purchase less coverage. I find strong support of the managerial signaling hypothesis.
7. Finally, if D&O insurance is an entrenchment device for corporate boards that do not want to assume the impact of their decisions, D&O insurance coverage and managerial entrenchment should be positively linked. The evidence concerning the entrenchment hypothesis is mixed.

² In Canada, Bill 198 of the Ontario Securities Act introduced the notion of derivatives lawsuits into Canada [see Emerson and Clarke (2003), LaCroix (2012), and Heys and Berenblut (2012)].
The next section of the paper presents the D&O insurance market and previous research on the topic. Section 3 presents the data and the testable hypotheses used to explain D&O insurance demand. In particular, I present the shareholder protection hypothesis, which is the foremost contribution of this paper. Section 4 presents the results. The results section is divided into three parts. The first part determines the likelihood that an agent will purchase D&O insurance. The second part finds the determinants of the policy limit and deductible amount. I conduct robustness checks in the last part of section 4. In section 5, I discuss my results and conclude.

2. Directors’ and officers’ liability insurance

2.1 The D&O insurance market

Contrary to the late eighties, the D&O insurance market was very soft (i.e., high coverage, low premium) in the mid to late 1990s. The liability insurance crisis of the mid 1980s [see Winter (1991)] had reduced the industry's capacity to provide insurance. This supply shock had a direct impact on premiums paid, which increased tenfold between 1984 and 1988, according to the standardized premium index for D&O insurance [see Larsen (2000)]. Feldhaus and Klein (2001) report that entry and exit in the professional liability insurance market is easy so that the D&O insurance market is quite competitive.

A D&O insurance policy comes to life when a manager is being sued as a representative of the corporation. The insurance company indemnifies the corporation and/or the manager only if the manager acted in good faith on behalf of the company; i.e., managers are not covered in case of gross negligence or criminal behavior. Depending on the type of D&O contract (there are more than 40 standard D&O insurance contracts), sometimes the manager will be indemnified directly, other times the corporation will be indemnified for the expenses incurred in the manager’s defense (for example, when the corporation has its own in-house legal team). As with traditional insurance contracts, D&O insurance contracts stipulate a premium to be paid, a policy limit as well as a deductible.

Most D&O insurance contracts are written on a claims-made and reported basis (CMR). CMR contracts differ from regular occurrence-based contracts in that they cover losses that are made and reported during the policy year even though such claims may have been incurred in previous years.

Occurrence-based contracts cover losses that are incurred during the policy year, no matter when the claim is reported in the future.³

So-called sticky points also characterize the D&O insurance coverage market. Typically, coverage limit is sold by layers of $1m, although the most important steps appear to be $5m. For our sample of companies, out of the 173 that purchased D&O insurance in any given year, two-thirds chose one of six policy limits: $5m, $10m, $15m, $20m, $25m and $50m.

2.2 Previous literature on D&O insurance

The first studies on D&O insurance using public data occurred in Canada and the U.K. [see Core (1997, 2000) and O’Sullivan (1997)]. Using a sample of 222 firms whose fiscal year ended between 31 May 1994 and 31 December 1994, Core (1997) finds that the most important determinants of D&O insurance purchase are the risk of a lawsuit and the cost of financial distress. No evidence is shown that D&O insurance is in any way part of the managers’ compensation package. In a follow-up article, Core (2000) finds that the factors explaining premiums are about the same as the factors explaining the demand for D&O insurance. Core's results are supported by O’Sullivan (1997) who concludes that in the U.K., D&O insurance coverage and managerial share ownership are corporate governance instrument substitutes. Moreover, Holderness (1990), Core (2000) and O’Sullivan (1997) suggest that D&O insurance acts as a monitoring device [see also Kim (2004), and Kaltchev (2006)].

If D&O insurance reduces the expected cost of bankruptcy and acts as a corporate governance instrument or monitoring device, one should expect stock returns to be positively correlated with D&O insurance purchases. Surprisingly, however, D&O insurance does not seem to have any impact on stock returns in the short run [see Bhagat et al. (1987), and Janjigian and Bolster (1990)].

³ For example, suppose that an incident occurs in 2001 (say the accidental pollution of a river), but is not reported until 2005 when a claim is filed. Under a CMR contract, all the financial responsibility for the loss falls upon the 2000 insurer. Under an occurrence-based insurance contract, it is the 2005 insurer that is responsible. See Doherty (1991), and Boyer and Gobert (2011) for more details regarding claims made and reported insurance contracts compared with occurrence-based contracts.
Chalmers et al. (2002) study the interaction between D&O insurance coverage and initial public offering (IPO) underpricing using an original sample of firms going public with a specific investment bank. Their study, using 72 observation points, is based on earlier studies by Ibbotson (1975) and Tinic (1988) who argued that IPOs are initially underpriced as a way to prevent shareholder lawsuits when the stock falls in the long run. They find that corporations with substantial D&O insurance coverage were, on average, more likely to be sued in the future for mispricing. In other words, the greater is the D&O insurance coverage, the less underpriced is the stock at the time of the IPO [see Boyer and Stern (2013), for a counter point].

Only a few papers examine in detail the modeling of D&O insurance. Gutierrez (2003) presents a model where D&O insurance is used as a commitment device for stockholders to monitor the behavior of corporate directors. The paper proposes a model where an uninformed principal purchases insurance on behalf of an informed agent because the insurer cannot observe ex-post the true effort of the agent. As a result, D&O insurance protects shareholders against the risk that corporate directors will not have enough wealth to pay for the damages they have caused. Consequently, D&O insurance acts as a deep-pocket, last-chance payer for shareholders who suffered a loss because of managerial incompetence. Parry and Parry (1991), Barney (1992), Kremslehner (2011) and Chang and Yeh (2011) offer other theoretical rationale for the existence of D&O insurance.4

If D&O insurance acts an out-of-the-money put option for the shareholders, would it not be cheaper to hold such put options in the company's treasury? Regulation or fear of litigation may prevent the company from doing exactly that [see for example Chacko et al. (2001)], so that corporations must find a contrived way to protect their shareholders from managerial accidents. Moreover, in the event of bankruptcy and liquidation, shareholders would not benefit from these put options since they would be seized by the firm's creditors as part of the firm's assets.

3. Hypotheses, data collection and data source

Although there are many hypotheses that one can test to find the determinants of the demand for D&O insurance, the main hypothesis I want to test in this paper is the one I refer to as the shareholder protection hypothesis. In the discussion, I refer to the protection or the coverage associated with a given D&O insurance policy; more protection (or more coverage) means that the policy limit is larger or that the deductible is smaller or both.

3.1 Main hypothesis: shareholder protection

Since shareholders are those that have the most to lose because of managerial incompetence or error, the shareholder protection hypothesis claims that it is them who should value D&O insurance protection the most. In other words, the more shareholders have to lose because of the managers' incompetence, the more protection they will purchase.

The market value of equity gives us the maximum possible loss that can befall shareholders. As the size of the possible loss increases, more protection becomes needed. To see why, consider the case of a corporation where all the board members are homeless. If the board becomes liable because of an error on their part, shareholders may not be compensated justly for their loss, as the directors have no personal wealth to be seized. D&O insurance then allows shareholders to claim back part of the wealth they lost. If the shareholder protection hypothesis is correct, we should see a positive relationship between a corporation's market value of equity and the amount of D&O insurance protection. Because we are dealing with large dollar values, I shall use the log of the market value of equity in millions of Canadian dollars5 (LnMVE) as my measure for the shareholders' wealth.

Another way to measure whether D&O insurance protects the firm's providers of capital rather than its directors would be to look at the use of debt in the corporation's financial structure. As the level of debt increases, it becomes less likely that any residual assets will be left to distribute to shareholders in the event of liquidation. Shareholders should then value a tool that compensates them in the event of liquidation due to mismanagement. I shall use the debt ratio, which is calculated as

---


5 All monetary figures are denominated in Canadian dollars; any amount in U.S. dollar has been converted to Canadian dollars using the exchange rate at year-end.
the ratio of the book value of debt to the quasi-market value of assets, as my measure of debt holder wealth in the corporation:

\[
\text{DebtRatio} = \frac{\text{Book value of debt}}{\text{Book value of debt} + \text{Market value of equity}}
\]

If shareholders use D&O insurance to protect themselves, we should see a positive relationship between D&O insurance coverage and the use of debt in the corporation.

3.2 Other hypotheses and determinants
Many other hypotheses have been developed to explain the corporate demand for risk management and insurance. I test for the most common.

3.2.1 Compensation package hypothesis
Core (1997) argues that D&O insurance is part of the director’s compensation package so that directors who have better coverage should be willing to accept a smaller compensation. I, therefore, need to control for the directors’ compensation. To do so, I shall use three measures for board members’ annual compensation: cash, options and stock. According to the compensation package hypothesis, D&O insurance coverage and director compensation should be negatively correlated.

Average annual cash compensation (in thousands of dollars) received by each board member is computed as total cash compensation of the board divided by the number of board members. Similarly, the average option compensation (in thousands of dollars) is calculated as the number of options per board members multiplied by the stock volatility and the stock price at the end of the year. The stock variable gives me a measure of each board member’s involvement in the corporation. Similar to Tufano’s (1996) CEO’s wealth involvement in the corporation, I measure stock as the one-year variation in the average board member’s wealth in shares of the corporation.  

3.2.2 Risk aversion hypothesis
Because of the prevalence of limited liability provisions and corporate indemnification plans, risk aversion can hardly be thought of as being a theoretically important factor. Risk aversion may play a role in two instances, however. First, if the company goes bankrupt the corporate indemnification plans may not have enough funds to compensate the directors for third-party lawsuits. Second, if the limited liability provision is set high enough, risk aversion may become an important determinant.

Of all the individuals who sit on the board, those that need the most convincing are the external directors [see Parry and Parry (1991)]. Officers who sit on boards, also known as inside directors, receive other forms of compensation and benefits from the corporation and as such do not require D&O insurance protection as much. Independent directors, on the other hand, request better and more complete protection to sit on boards. As a result, if the director risk aversion hypothesis is correct, I expect the number of independent directors (Independent) on a corporate board to be positively associated with a higher level of protection. I define a director to be independent if he is not an officer of the corporation, related to an insider or a former employee.

3.2.3 Financial distress/corporate insurance hypothesis
If one considers insurance as part of a corporation’s overall risk management portfolio, then the corporate insurance hypothesis is probably the one that has received the most attention in the finance literature. Size should be an important determinant of a corporation’s need for corporate insurance; larger corporations are better able to self-insure so they have less need for insurance [see Mayers and Smith (1982), and Smith and Stulz (1985)]. For example, large corporations may have an in-house litigation department that handles lawsuits. To measure the impact of size, we shall use the previously defined variable Ln(MVE).

The way in which a corporation is financed should affect the need for corporate insurance. For example, a corporation near

---

6 This option value measure is unusual, but traditional option value measures are not perfectly computable using the data available because management proxies in Canada do not provide strike prices or expiration dates for the options granted to board members. Running the analysis using the average number of options granted or a value using a the Black-Scholes formula with a strike price equal to the stock’s year-end price and a maturity date of 10 years does not change the results. The same can be said about the use of a dummy variable that measures whether board members received stock options or not.

7 The value of the stock variable may change because there is change in the price of the stock or a change in the average number of shares owned by board members. Measuring only the ownership change does not alter the results.
bankruptcy will incur large financial distress costs. Moreover, since a bankrupt firm cannot honor its promise to financially support its managers' legal fees in the event of a lawsuit, the corporate indemnification plan becomes moot. As a result, corporate managers will require better protection to sit on boards of corporations near bankruptcy. Thus, bankruptcy risk should increase the need for D&O insurance coverage. Bankruptcy is calculated as in Core (1997) so that it measures the value of a put option on the corporation's assets with a strike price equal to the book value of debt. Financial distress also increases when the size of the corporate debt is relatively large. Corporations that have a high debt ratio (as defined earlier as our DebtRatio variable) are more at risk of being in financial distress.

Finally, a firm's past return may be a good indication of its financial health. As a consequence, the higher the return the more liquidity the corporation has to face a small market downturn. The need for D&O insurance should, therefore, be smaller when past returns have been high. I shall use the firm's return on assets (RoA) as my profitability measure. I calculate RoA as the ratio of net earnings to the book value of assets.

3.2.4 Monitoring and the entrenchment hypotheses

Holderness (1990), O'Sullivan (1997) and Core (2000) hypothesize that D&O insurance is a substitute for board monitoring. There are three other important ways to monitor the board: block holders, debt holders and independent directors. Although each measure is not independent of each other (see Agrawal and Knoeber (1996)) we should expect a negative relationship between them and D&O insurance coverage if the monitoring hypothesis is correct. Moreover, if D&O insurance is used to protect entrenched managers, we should again see a negative relationship between the presence of outside board monitors and D&O insurance coverage.

I shall use three measures of board monitoring devices: the presence of creditors, the presence of financial institution block holders and the presence of independent directors. Although each measure is not independent of each other (see Agrawal and Knoeber (1996)) we should expect a negative relationship between them and D&O insurance coverage if the monitoring hypothesis is correct. Moreover, if D&O insurance is used to protect entrenched managers, we should again see a negative relationship between the presence of outside board monitors and D&O insurance coverage.

Regarding the Blockholder variable, although one could make the argument that any block ownership may be a substitute for D&O insurance monitoring, not every block holder will have incentives to exert influence on the board. Presumably, financial institution block holders are those who are more likely to be substitute for D&O insurer monitoring. I measure Blockholder as the percentage of the shares held by financial institutions. Finally, the more independent directors on the board, the less D&O insurance monitoring is needed, so that the previously defined Independent variable should be negatively related to D&O insurance coverage.

According to Core (1997) and Chalmers et al. (2002), entrenched managers may use their power on the board to purchase more D&O insurance to extract wealth from shareholders without too much fear of reprisal. I use two main measures of entrenchment: the voting power of the CEO (CEOPower), and the wealth of insiders (InWealth) on the corporate board. In both cases, I expect a positive relationship if the entrenchment hypothesis is correct.

Because Canadian security laws allow the existence of different classes of voting shares, CEOPower is measured as the proportion of the votes in the hands of the CEO divided by the total number of votes in the corporation. InWealth is calculated as the total number of shares held by insiders divided by the total number of shares. I define a corporate insider as any corporate officer or employee who sits on the board in addition to the chairman of the board.

3.2.5 Signaling hypothesis

The adverse selection literature [see Rothschild and Stiglitz (1976)] predicts that good managers will signal their quality by having less insurance, although DeMarzo and Duffie (1991) and Bessembinder (1991) argue that good managers will signal their quality by hedging idiosyncratic risk more. Moreover, signaling is even more important when managerial quality is harder to assess because their tasks are complex. I shall use two measures of complexity. First, managers of growth corporations have complex investment opportunities so that they may benefit more from signaling their quality than managers in other types of corporations. As a result, growth corporations should purchase lower D&O insurance coverage. I measure Growth as the ratio of the market value of equity to its book value. A second measure of complexity is the corporation's stock return volatility (Volatility). Managers in corporation whose stock is very volatile are more
likely to signal their quality through lower insurance coverage. Lower coverage is thus expected when the stock is volatile. In insurance, signaling is usually done through the use of higher deductibles rather than smaller policy limit.

3.2.6 Litigation risk and other control variables
One last determinant of D&O insurance demand is the perceived litigation risk a corporation faces. Through strategic corporate choices, a firm may increase its risk of being sued even though corporate managers have done nothing wrong. Sometimes, the environment is simply more litigious. For example, a Canadian corporation that has important activities in the United States should be faced with more litigation than a corporation whose main activities are in Canada.

As reported by Core (1997) and Clarkson and Simunic (1994), the Canadian legal system is less litigious than the American legal system. It follows that claims filed in the United States should be costlier than claims filed in Canada. Given that almost half of the D&O lawsuits originate from shareholders, I feel that the most important litigation risk that corporations face is whether they have violated some SEC rule, typically an accounting irregularity. The Ontario Securities Commission introduced (and effective in 2005) the notion of derivatives lawsuits into Canada [see Emerson and Clarke (2003)]. It follows that corporations that are listed on a stock exchange in the U.S. (U.S.Listed) should have a greater policy limit and a larger deductible.

Another example, mentioned by Core (1997), is that corporations are the most likely to be sued if they engage in mergers, acquisitions and divestiture, so that increased litigation risk should be a function of large variations in the corporation’s book value of assets. I let Acquiror equal the log of the book value of assets (in millions of dollars) if the said book value has increased by more than 25% over the previous year. Divestor, on the other hand, equals the previous year’s log of the book value of assets (in millions of dollars) if the book value has decreased by more than 25% in the current year. An increase in the litigation risk increases the frequency and the severity of lawsuits, so that we should expect an increase in the deductible and in the policy limit.

The last variable I need to include in the analysis takes into account the length of the D&O insurance contract. It was common practice in the nineties for insurers to issue multiyear contracts with automatic reinstatement clauses. To control for these multiyear insurance contracts, I count the number of years during which the parameters of the contract (deductible and policy limit) did not change. Because it appears that no contract is ever longer than three years, the MultiYear variable may only be equal to 1, 2 or 3.

The astute reader will note that there is no mention of the convexity of the tax schedule in the present analysis, even though progressive tax systems induce corporations to smooth earnings [see Smith and Stulz (1985), and Graham and Smith (1999)]. The reason is that D&O insurance offers little tax advantages. Consequently, there is no reason why the convexity of the tax structure should have any bearing on the demand for D&O insurance. Moreover, Graham and Rogers (2002) find that corporations do not seem to alter their hedging strategy as a function of the convexity of the schedule, which reduces the strength of the general tax argument for purchasing insurance.

3.3 Summary of hypotheses and predicted impact
Table 1 presents a summary of the predicted impact of each variable as a function of the hypothesis it claims to test. The + (−) sign means that the variable in the left column is expected to have a positive (negative) impact on the need to carry D&O insurance coverage. For example, for LnMVE, the financial distress hypothesis predicts that larger corporations need less insurance (thus the − sign) because they face smaller financial distress costs and are better able to self-insure. On the other hand, the shareholder protection hypothesis predicts that larger corporations will choose more protection (thus the + sign) because shareholders have more to lose.

The independent variables I use to test my hypotheses have also been used to test other hypotheses. For example, Tufano (1996) uses the managers’ stock and stock option ownership to test for their managerial risk aversion. He suggests that stock ownership increase the managers’ implicit risk aversion whereas
stock option ownership decreases it. This means that, under the Aversion column in Table 1, we could put a + next to the stock variable and a − next to the option variable. Also, DeMarzo and Duffie (1991) and Bessembinder (1991) suggest that more able managers should use hedging the most because it eliminates the noise component in their decisions so that all that remains is their quality.

The choice I made in presenting the different hypotheses and the measures used to test them are subject to debate. However, since the main goal of the paper is to test whether D&O insurance protects the shareholders more than the managers, any other hypothesis test is secondary to the shareholder protection hypothesis.

### 3.4 Dependent variables

I use three different dependent variables. First, I use a dichotomous variable to see whether D&O insurance was purchased or not. Second, I have the policy limit that gives us the maximum possible indemnity paid by the D&O insurance policy for any single event, or for any single year. Although D&O insurance is usually sold in excess layers (see Chalmers et al. (2002)), I do not have access to that information. Third, I have information regarding the contracts’ deductibles. This deductible gives me the amount the corporation (or the managers) must pay out-of-pocket before the insurance coverage kicks in. To reduce problems associated with the extreme measures of policy limit and deductible, I used the logarithm of the dollar figure in my analysis. In the case of deductibles, there are times where the insurance contract specifies an individual deductible, a corporate deductible, or both. For the purpose of this paper, I used the highest amount available, typically the corporate deductible.

Although there are no standard D&O insurance policy forms, they all include a policy limit and a premium paid. The non-standard aspects of the contracts relate to the exclusions and inclusions, the length of the policy and who is covered. Unfortunately, I do not have access to the specific insurance contract, but only to what is reported in the information circular.

### 3.5 Data source

Financial data was obtained from three different sources (Compustat, Stock Guide and CanCorp Financial), which allows me
Directors’ and officers’ insurance and shareholder protection

The original sample included 354 Canadian corporations drawn from seven economic sectors: bio-pharmaceutical, forest and paper, industrial products, technological products, consumer products, merchandising, and media. These sectors were chosen based on the TSE sector list drawn from the TSE FactBook. Two important sectors of the Canadian economy (financial institutions, and mining and natural resources) were deliberately omitted to keep our sample more homogenous. The fiscal years used are from 1993 to 1998. For each economic sector, all corporations listed in the TSE publication and traded on the exchange were included in the dataset for each year of my study. The exercise yielded a total of 1519 observations.

Because of incongruities in the financial and management proxies (for example, board or CEO ownership of more than 100% of the company’s stock and no trading in the stock during the year), 112 observations were removed from the original dataset. My final dataset includes 1407 observations for 318 firms. This dataset presents no apparent survivor bias because each year we added firms to the data based on the TSE FactBook. Also, companies disappeared during the period under study, either because of a merger or bankruptcy. This incomplete rolling panel yields an average of 4.4 observations per firm. Of the 318 firms in the final sample, close to 60% have information for five or six years and 17% no longer existed at the start of 2000. Table 2 presents D&O insurance penetration for each sector and each year. Insurance penetration is calculated as the proportion of corporations that purchased D&O insurance. On average, 70% of the firms purchased D&O insurance, although penetration seems to increase over the years (67% in 1993, 73% in 1998).

### Table 2: D&O insurance penetration by economic sector and fiscal year

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-pharmaceutical</td>
<td>75%</td>
<td>88%</td>
<td>92%</td>
<td>92%</td>
<td>90%</td>
<td>90%</td>
<td>89%</td>
</tr>
<tr>
<td>Forest and paper</td>
<td>59%</td>
<td>58%</td>
<td>63%</td>
<td>64%</td>
<td>67%</td>
<td>70%</td>
<td>70%</td>
</tr>
<tr>
<td>Industrial products</td>
<td>68%</td>
<td>69%</td>
<td>71%</td>
<td>76%</td>
<td>74%</td>
<td>74%</td>
<td>64%</td>
</tr>
<tr>
<td>Technological products</td>
<td>100%</td>
<td>90%</td>
<td>90%</td>
<td>84%</td>
<td>70%</td>
<td>68%</td>
<td>74%</td>
</tr>
<tr>
<td>Consumer products</td>
<td>68%</td>
<td>65%</td>
<td>70%</td>
<td>67%</td>
<td>63%</td>
<td>67%</td>
<td>67%</td>
</tr>
<tr>
<td>Merchandising</td>
<td>52%</td>
<td>57%</td>
<td>58%</td>
<td>56%</td>
<td>55%</td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>Media</td>
<td>77%</td>
<td>78%</td>
<td>78%</td>
<td>73%</td>
<td>79%</td>
<td>85%</td>
<td>85%</td>
</tr>
<tr>
<td>Weighted average</td>
<td>83%</td>
<td>67%</td>
<td>69%</td>
<td>71%</td>
<td>70%</td>
<td>71%</td>
<td>73%</td>
</tr>
</tbody>
</table>

Percentage of corporations by economic sector and by year that carry D&O insurance.
which is slightly higher than the 63% penetration reported by Core (1997). Although Core relies on the same public information source as mine, his sample of companies is different because he includes public utility, financial and mining corporations.

Of the 1,407 firm-years, 988 report having D&O insurance. Of these, I have information about the policy limit for 985 firm-years; the other 124 cases are missing. This causes an identification problem: was the deductible omitted by mistake or was it equal to zero and not mentioned as such? Three possibilities are offered to me. I can set the deductible to zero, I can eliminate the 124 observations for which no deductible is mentioned or I can set the deductible to the lowest amount in our dataset ($10,000). I opted for the latter, although the results are sensibly the same whichever approach I use.

4. Methodology and results
An important methodological assumption I make in the present analysis is that the insurance company and the board of directors know what the financial health of the corporation is before negotiating over the D&O insurance contract. In other words, the annual report is known before the D&O insurance contract is chosen. Moreover, given that the cost to a firm of purchasing a D&O insurance policy is very small compared to its total cash flows (an average premium of $85,000 versus assets of $1b and total debt of over $700m), I feel it is reasonable to assume that the demand for D&O insurance does not affect the firm's financial and operational decisions. As a consequence, I shall treat all the information related to each firm's financial status (leverage, bankruptcy risk, stock volatility, etc.) as being exogenous when determining the choice of D&O insurance coverage.

The current analysis focuses only on the demand side. The entire supply side of the D&O insurance market is untouched. The underlying logic is that the D&O insurance market is very competitive, especially during the nineties. Moreover, as Core (1997) reports, "the vast majority of firms could obtain insurance at some price, no matter how risky they were" (page 66). As a result, any firm may purchase as high a policy limit as it requires [see also Doyle (1991), and Feldhaus and Klein (2001), for more details on the competitiveness of the D&O insurance market].

4.1 D&O purchase: comparing means and likelihood

4.1.1 Descriptive statistics
Table 3 presents the variables' summary statistics after removing observations for which the lagged value of assets is not available. The number of observations is thus reduced from 1407 to 1074, which means that the number of firm-years where D&O insurance was purchased is reduced from 988 to 770. The mean of every variable is calculated for corporations that carried D&O insurance (first set of numbers) and for corporations that did not carry D&O insurance (second set of numbers). It appears that corporations that purchase D&O insurance are larger (LnMVE), are more likely to be listed in the U.S. (U.S.Listed) and more likely to have engaged in a round of mergers and acquisition in the past year (Acquiror). Firms that carry D&O insurance are only marginally more likely to compensate their directors through the use of stock options. The reason may be that very few corporations distribute options to their directors: only 14% of D&O insurance purchasers and 11% of non-purchasers give stock options to their managers. Corporations that carry D&O insurance also appear to be more growth oriented, but only marginally so. Finally, corporations that carry D&O insurance have a smaller debt ratio than corporations that do not.

4.1.2 Likelihood of carrying D&O insurance
Before exploring what explains D&O insurance policy limits and deductibles, it would be interesting to find the determinant of the purchase, or not, of the policy. Once I have determined what type of corporations purchase D&O insurance, I will then be able to study what determines the policy limit and the deductible, given that the corporation has purchased the insurance policy.

Looking back at Table 3, one could expect that size, financial structure, growth opportunities and listing on an American stock exchange could be important determinants of D&O insurance purchases. When I run a probit regression, using all the independent variables, only four variables appear to have any significance: Bankruptcy, DebtRatio, U.S.Listed and Acquiror. Except for the debt ratio variable that reduces the likelihood of carrying D&O insurance, the other three variables appear to

10 Table not shown. It is important to mention that this regression does not use a dummy variable for the year or for the firm.
increase its likelihood. This analysis does not take into account firm effects or time effects. Because my panel data has many observations for the same corporation, the draws are not independent from year to year.

Interestingly, during the time period under study, corporations either purchased D&O insurance all the time or never purchased D&O insurance. This means that a simple firm fixed effect explains all the decision to purchase D&O insurance. As a result, none of the aforementioned hypotheses can be tested based on whether the company purchased D&O insurance or not. This may be due to the managers’ unwillingness to change their approach to insurance and risk management by fear of appearing weak; this would be tantamount to confessing to have made a mistake, as shown in Prendergast and Stole (1996), Dewatripont and Tirole (1996), and Boyer and Moreaux (1997). As a result, the reasons why some corporations never hedge have nothing to do with efficient risk management strategies and positioning; instead, it has to do with the managers’ belief that hedging idiosyncratic risk is not useful or even detrimental to the firm.

The fact that D&O insurance demand appears to be a matter of belief is consistent with the industry view that denied D&O insurance coverage is almost unheard of [Chalmers et al. (2002)]. Moreover, the structure of the D&O insurance contract as a claims-made and reported policy may explain why corporations that purchase D&O insurance do so for a long time [Boyer and Gobert (2011)]. It is still problematic to explain why directors in

<table>
<thead>
<tr>
<th>Variable</th>
<th>Obs.</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Obs.</th>
<th>Mean</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ln (Limit)</td>
<td>770</td>
<td>16.61</td>
<td>(1.00)</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Ln (Deductible)</td>
<td>770</td>
<td>11.20</td>
<td>(1.45)</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Longterm</td>
<td>770</td>
<td>1.86</td>
<td>(0.83)</td>
<td>NA</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>Cash</td>
<td>770</td>
<td>5.57</td>
<td>(10.24)</td>
<td>304</td>
<td>5.79</td>
<td>(14.91)</td>
</tr>
<tr>
<td>Option</td>
<td>756</td>
<td>0.064</td>
<td>(0.915)</td>
<td>301</td>
<td>0.004</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Stock</td>
<td>764</td>
<td>4.29</td>
<td>(51.98)</td>
<td>304</td>
<td>3.19</td>
<td>(25.32)</td>
</tr>
<tr>
<td>Independent</td>
<td>770</td>
<td>5.50</td>
<td>(2.38)</td>
<td>304</td>
<td>5.37</td>
<td>(2.81)</td>
</tr>
<tr>
<td>LnMVE</td>
<td>764</td>
<td>5.16</td>
<td>(1.87)</td>
<td>304</td>
<td>4.82</td>
<td>(1.94)</td>
</tr>
<tr>
<td>Bankruptcy</td>
<td>730</td>
<td>-6.60</td>
<td>(6.57)</td>
<td>301</td>
<td>-6.63</td>
<td>(6.38)</td>
</tr>
<tr>
<td>DebtRatio</td>
<td>764</td>
<td>0.402</td>
<td>(0.237)</td>
<td>304</td>
<td>0.438</td>
<td>(0.217)</td>
</tr>
<tr>
<td>RoA</td>
<td>770</td>
<td>0.013</td>
<td>(0.145)</td>
<td>304</td>
<td>0.021</td>
<td>(0.183)</td>
</tr>
<tr>
<td>Blockholder</td>
<td>766</td>
<td>5.50</td>
<td>(13.10)</td>
<td>303</td>
<td>5.70</td>
<td>(13.43)</td>
</tr>
<tr>
<td>Growth</td>
<td>764</td>
<td>13.16</td>
<td>(158)</td>
<td>304</td>
<td>2.12</td>
<td>(4.09)</td>
</tr>
<tr>
<td>Volatility</td>
<td>735</td>
<td>0.506</td>
<td>(0.285)</td>
<td>301</td>
<td>0.530</td>
<td>(0.348)</td>
</tr>
<tr>
<td>CEOPower</td>
<td>770</td>
<td>0.122</td>
<td>(0.206)</td>
<td>304</td>
<td>0.137</td>
<td>(0.233)</td>
</tr>
<tr>
<td>LnWealth</td>
<td>764</td>
<td>169</td>
<td>(1301)</td>
<td>304</td>
<td>161</td>
<td>(1156)</td>
</tr>
<tr>
<td>U.S.Listed</td>
<td>770</td>
<td>0.151</td>
<td>(0.358)</td>
<td>304</td>
<td>0.095</td>
<td>(0.294)</td>
</tr>
<tr>
<td>Acquiror</td>
<td>770</td>
<td>1.58</td>
<td>(2.59)</td>
<td>304</td>
<td>1.14</td>
<td>(2.27)</td>
</tr>
<tr>
<td>Divestor</td>
<td>770</td>
<td>0.189</td>
<td>(0.908)</td>
<td>304</td>
<td>0.153</td>
<td>(0.867)</td>
</tr>
</tbody>
</table>

Table 3: Descriptive statistics for the 1074 usable observations by insurance purchasing or not

Because of the lagged variable used in the analysis, the number of observations has decreased from 1407 to 1074. By definition the limit, the deductible and the contract length are not listed for corporations that did not report to have D&O insurance. The * (#) represents variables whose mean is statistically different at the 5% (10%) level between firms that purchased D&O insurance and firms that did not. Obs. is the number of observations, and St. Dev. is the standard deviation.
corporations that do not purchase D&O insurance do not care about it.

4.2 Policy limit and deductible

The Probit analysis suggests that a corporation’s financial health and structure, and its board’s compensation and composition have no impact on its decision to purchase D&O insurance or not. Nothing is said in those results about the D&O insurance coverage in firms that have decided to purchase some. Firm-specific conditions may have an impact on the chosen policy limit or deductible. This section of the paper deals with the demand for D&O insurance, given that it was purchased.

In studies on insurance demand and hedging, it is normally important to control for the fact that the dependent variable is truncated at zero because we do not know the policy limit choice of corporations that did not purchase any insurance. As a result, standard econometric methodology (Green (1997)) asks us to use a Tobit regression approach and the Mills’ ratio to control for those unobservable. This is the approach used in the Core (1997) study.

A Tobit regression is not warranted here, however, because the decision to purchase D&O insurance has nothing to do with any financial, corporate governance or litigation risk measure. It is rather based on whether the corporation purchased D&O insurance the previous year and on some fixed firm characteristic. Thus, not only can I discard completely from the analysis all firms that never purchased D&O insurance, but I do not even need to control for the selection bias.

Table 4 presents the results of an ordinary least square regression with fixed effect with all the independent variables. The first regression model in Table 4 explains the policy limit choice and the second the deductible choice. The regression model assumes that the deductible and the policy limit are chosen independently of each other. This may not be the case, however. Policyholders usually choose the policy limit before they negotiate the deductible. It may then be important to control for the timing in the choice variables, which I shall do in Section 4.3 of the paper. The fixed effect regression models are able to explain more than 90% of the variations in both dependent variables.

4.2.1 Policy limit

In the policy limit regression, only director cash compensation, market value of equity, debt ratio and the presence of large financial institution block holders appear to have a significant impact on the policy limit. The negative sign on cash supports the efficient compensation package hypothesis, even though the two other variables used to measure this hypothesis (option and stock) are not significant. The insignificant impact of option and stock may be due to the fact that very few corporations compensate their outside directors through stock and stock option grants. Moreover, because litigation occurs when the stock price falls precipitously, managers may feel that compensation linked to the value of the corporation is not a good substitute for D&O insurance coverage. The negative sign on cash suggests that board members consider D&O insurance as part of their overall compensation package. As a result, they are willing to sacrifice part of their cash compensation for protection against adverse negative shocks to their personal wealth in the event of a lawsuit.

With respect to the shareholder protection hypothesis, strong support is shown through the combination of the positive sign on both the DebtRatio variable and, more importantly, on the LnMVE variable. The two variables used to measure this hypothesis are significant at the 1% level in explaining the D&O insurance policy limit. Thus, when shareholders have more wealth to lose and/or when creditors are relatively more involved, corporations purchase more coverage against managerial incompetence. Because this insurance coverage benefits disproportionately the shareholders, as suggested by Romano (1991a), the shareholder protection hypothesis appears validated by the empirical analysis. Put in the context of homeowner insurance, as the value of the house (the asset at risk) increases, individuals (the shareholders and the creditors) want to protect the value in their house through an increase in insurance coverage (typically an increase in the policy limit).

It is true that the positive sign on DebtRatio lends support to both the shareholder protection and the financial distress hypotheses. The support for the latter is, however, weaker than for the former for two reasons. First and foremost, the positive sign on LnMVE contradicts the financial distress hypothesis, although it confirms the shareholder protection hypothesis. Second, no other variable used to measure the cost of financial distress (bankruptcy and RoA) is significant.
Another hypothesis for which I only find weak support is the monitoring hypothesis that states that D&O insurance coverage was a substitute for other forms of managerial control such as the presence of creditors and of large block holders. I, therefore, expected a negative relationship between coverage and the DebtRatio and Blockholder variables. With respect to Blockholder, the monitoring hypothesis finds support, but the impact of DebtRatio on coverage is opposite to what was expected. Consequently, and similarly to the financial distress hypothesis, support for the monitoring hypothesis is not convincing because: (i) only the Blockholder variable supports it and (ii) another measure used to test it, namely DebtRatio, has the wrong sign (and significantly so).

With respect to the risk aversion, managerial quality signaling and managerial entrenchment hypotheses, I find no significant impact on the policy limit choice. Even the variables used to control for higher litigation risk, including whether the corporation is listed in the U.S., do not seem to have an impact on the D&O insurance policy limit.

Table 4: The determinants of policy limit and deductible amounts ordinary least square regressions with firm and year fixed effects

The dependent variables are the log of the policy limit and the log of the deductible. Ordinary least square regression using time and firm fixed effects. The hypothesis or hypotheses that each independent variable wants to test is presented in the second column. Also, the second column gives the predicted sign of each coefficient on the dependent variable; the first sign is the predicted impact on the policy limit and the second sign is the predicted impact on the deductible. The ** represents a coefficient that is significant at the 1% level, * significant at the 5% level and # at the 10% level. Some coefficients have been scaled for presentation purposes. Standard errors are in parentheses. N=696.
4.2.2 Deductible
In the deductible regression, it is interesting to see that none of the variables that had any explanation power in determining the limit have any power in determining the deductible. Rather, the four variables that have any explanatory power are the measure of growth opportunity, stock volatility, corporate insider wealth and whether the corporation is listed in the U.S. The first three variables suggest that the managerial signaling and the managerial entrenchment hypotheses are both supported by the data.

The signaling hypothesis predicted that when it is harder to assess the quality of a manager, then good managers will want to signal their quality by choosing an insurance contract that protects them less. In other words, they will choose an insurance contract with a higher deductible. In my analysis, I used the presence of growth opportunities and the volatility of market returns as instances when managers are more likely to want to signal their quality. In the regression, both the Growth and the Volatility variables are positively correlated with the deductible, just as the managerial signaling hypothesis predicted. This result is in accordance with standard adverse selection in insurance results where agents prefer to signal their quality through a higher deductible rather than a lower limit. The case for the managerial signaling hypothesis is, therefore, strong in the determination of the deductible, as it was for the shareholder protection hypothesis in the determination of the policy limit.

With respect to the managerial entrenchment hypothesis, it appears that the more wealth insiders have in the corporation, the more protection the corporation purchases on their behalf through a lower deductible.

Similar to the case of the policy limit, I find no evidence that managerial risk aversion plays any role in the determination of the deductible. This may suggest that limited liability plans and corporate indemnification provisions protect directors sufficiently so that they do not need D&O insurance as much.

4.2.3 Discussion
It is interesting to see that no hypothesis is supported by the policy limit and the deductible regressions simultaneously. This raises questions related to the possibility of model over-fitting. Another question raised by the results in Table 4 is that the motivation behind the determinants of the policy limit and of the deductible are not the same so that different hypotheses explain different aspects of the demand for D&O insurance. To see whether I am over-fitting the model, I restrict the regression model to include only the variables that have predictive power on each dependent variable. The regression models are thus

\[ \text{Deductible}_{i,t} = \text{Growth}_{i,t} + \text{Volatility}_{i,t} + \text{LnWealth}_{i,t} + \text{U.S.Listed}_{i,t} + \epsilon_{i,t} \]

The regressions results are displayed in Table A1 in the Appendix. Put succinctly, all the variables used in the restricted model have an impact on the corresponding dependent variables that is the same as in Table 4. Thus, there does not appear to be any over-fitting problem associated with my main results.

The only difference between Panel A and B in Table A1 is that more observations are used in Panel B. Why is that? Note that in the restricted model, no variable is constructed using a lagged observation (i.e., Acquiror and Divestor are not used). As a result, I am able to add observations that were deleted in the regressions in Table 4. The regression results are sensibly the same; the only difference being that Volatility is no longer significant in explaining the deductible. The reduction in significance is not very large, however, since Volatility was only significant at the 10% level in Table 4.

These ordinary least squares (OLS) results do not consider the fact that, usually, policyholders choose the policy limit before the deductible so that, when the time comes to choose the deductible amount, the policy limit is known. As a consequence the OLS results for the deductible may be biased. Another possibility is that the deductible and the policy limit are chosen simultaneously. I conduct these robustness checks in the next section.

4.3 Robustness
In this section of the paper, I test the robustness of the results in two ways. First, I control for possible endogeneity issues that arise from the choice of the policy limit and the deductible. Second, because D&O insurance limits are characterized with sticky points (see Section 2.1), the dependent variable is not continuous. Thus, an ordered Probit approach may be more appropriate.
4.3.1 Endogeneity
Endogeneity may take on many forms. I test for two, which are presented in Tables A2 and A3 in the Appendix. First, I see whether the sequence of choice in the insurance contract has any impact on the dependent variables. Second, I test for simultaneity in the choice of each dependent variable to see whether the independent assumptions are not driving the results.

The first robustness check I conduct is to test whether the deductible is chosen after the limit during the negotiations between the corporation and the insurance broker. Because it is through a higher deductible that a policyholder can most reduce his premium, and because the policy limit is the most important part of the insurance contract, it makes intuitive sense to negotiate the policy limit first, then to negotiate the deductible. As a result, the policy limit is known when time comes to choose the deductible. The deductible regression is presented in Table A2. By adding the policy limit as a determinant of the deductible does not change the results greatly compared to those presented in Table 4.

A second source of endogeneity may be that the policy limit and the deductible are chosen simultaneously. To test for this endogeneity, I shall use a two-stage instrumental variable regression of the form

$$\text{Limit}_{i,t} = \text{Cash}_{i,t} + \text{InMVE}_{i,t} + \text{DebtRatio}_{i,t} + \text{Blockholder}_{i,t} + \text{Longterm}_{i,t} + \text{Deductible}_{i,t} + d_t + d_i + \epsilon_{i,t}$$

and

$$\text{Deductible}_{i,t} = \text{Growth}_{i,t} + \text{Volatility}_{i,t} + \text{InWealth}_{i,t} + \text{U.S.Listed}_{i,t} + \text{Limit}_{i,t} + d_t + d_i + \epsilon_{i,t}$$

Where $\text{Limit}_{i,t}$ and $\text{Deductible}_{i,t}$ represent the predicted value of each variable obtained from the reduced-form regressions. The reduced-form regressions are

$$\text{Limit}_{i,t} = (X_{i,t}^{\text{Limit}} \cup X_{i,t}^{\text{Deduct}}) + d_t + d_i + \epsilon_{i,t}$$

and

$$\text{Deductible}_{i,t} = (X_{i,t}^{\text{Limit}} \cup X_{i,t}^{\text{Deduct}}) + d_t + d_i + \epsilon_{i,t}$$

For notation, I let $X_i$ be the matrix of independent variables that explains each dependent variable, $d_t$ is the year dummy variable and $d_i$ is the firm dummy variable. $(X_{i,t}^{\text{limit}} \cup X_{i,t}^{\text{Deduct}})$ is the union of the independent variables.

The results from the structural form regressions are presented in Table A3 (the results from the reduced-form regressions are not shown). As we see, all variables remain significant and are of the anticipated sign, including the predicted dependent variables. The results are sensibly the same as those obtained in Table A1 when we ran a restricted OLS regression using firm- and time-fixed effects.

As in Panel A of Table A1, no variable used in Panel A of Table A3 relies on the use of a lagged observation. As a result, I am able to add some observations back to regression. The two-stage regression results from the restricted model using all available observations are presented in Panel B of Table A3. Overall, the regression results are basically the same as those presented in Panel B of Table A1, but with some variables having more explanatory power and others having less. I can, therefore, be reasonably certain that the endogeneity problem is not severe since the shareholder protection hypothesis is still supported even after controlling for potential endogeneity problems.

4.3.2 Probit
As mentioned earlier, the D&O insurance market is characterized by sticky points. As a result, the policy limit choice does not seem to follow a continuous distribution. In fact, six policy limit levels account for two-thirds of all D&O insurance policy sold in 1998: $5m, $10m, $15m, $20m, $25m and $50m. Other years have similar sticky points. It is then appropriate to take into account these jumps using an ordered Probit approach. The results from the ordered Probit are presented in Table 5. The results are sensibly the same as those presented in the main body of the paper.

The most important differences appear not in the determinants of the policy limit, but in the determinants of the deductible. In this Probit analysis, the risk of bankruptcy appears to reduce the size of the deductible in accordance with the financial distress hypothesis. At the same time, however, the debt ratio increases the deductible, thus invalidating the financial distress hypothesis. It is, therefore, not clear what conclusion to draw from these results. Similarly, the managerial entrenchment hypothesis is...
validated by the negative sign on the impact of insider wealth, but invalidated by the positive sign of the CEO voting power.

5. Discussion and conclusion

The goal of this paper was to offer a further insight into the corporate demand for insurance by focusing on one particular aspect: the demand for D&O liability insurance. Corporations purchase this insurance on behalf of their directors and officers to protect them in case a lawsuit is brought against them as representatives of the corporation. This paper tests many measures hypothesized to have an impact on the corporate demand for insurance. The most interesting result I find relates to the fact that directors’ and officers’ insurance offers protection not as much to D&O of the corporation, but rather to its shareholders. In fact, the shareholder protection hypothesis finds consistent support in all model specifications, including model specifications used in the Robustness section of the paper.

The study of D&O insurance remains a largely unexplored territory as few papers have been devoted to this aspect of corporate governance. Is D&O insurance part of the compensation package? Is it a tool to align the manager’s incentives with those of the shareholders? Is it truly designed to protect corporate directors or other stakeholders, like such as the shareholders?

### Table 5: The determinants of policy limit and deductible amounts ordered Probit analysis with firm and year fixed effects

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash</td>
<td>Package</td>
<td>-</td>
<td>+</td>
<td>-0.046**</td>
<td>(0.012)</td>
<td>-0.000</td>
<td>(0.013)</td>
</tr>
<tr>
<td>Option</td>
<td>Package</td>
<td>-</td>
<td>+</td>
<td>-0.086</td>
<td>(0.081)</td>
<td>-0.085</td>
<td>(0.061)</td>
</tr>
<tr>
<td>Stock</td>
<td>Package</td>
<td>-</td>
<td>+</td>
<td>0.000</td>
<td>(0.003)</td>
<td>-0.000</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Independent</td>
<td>Aversion</td>
<td>+</td>
<td>-</td>
<td>0.025</td>
<td>(0.057)</td>
<td>0.047</td>
<td>(0.056)</td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress shareholder</td>
<td>+</td>
<td>+</td>
<td>0.618**</td>
<td>(0.149)</td>
<td>0.151</td>
<td>(0.153)</td>
</tr>
<tr>
<td>Bankruptcy (%)</td>
<td>Distress</td>
<td>+</td>
<td>-</td>
<td>0.002</td>
<td>(0.011)</td>
<td>-0.019*</td>
<td>(0.012)</td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress monitor shareholder</td>
<td>+</td>
<td>+</td>
<td>2.041**</td>
<td>(0.698)</td>
<td>1.266*</td>
<td>(0.702)</td>
</tr>
<tr>
<td>RoA</td>
<td>Distress</td>
<td>-</td>
<td>+</td>
<td>-0.948</td>
<td>(0.662)</td>
<td>0.060</td>
<td>(0.614)</td>
</tr>
<tr>
<td>Blockholder</td>
<td>Monitor</td>
<td>-</td>
<td>+</td>
<td>-0.028**</td>
<td>(0.009)</td>
<td>-0.008</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Growth</td>
<td>Signaling</td>
<td>-</td>
<td>+</td>
<td>0.001</td>
<td>(0.003)</td>
<td>0.011*</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>-</td>
<td>+</td>
<td>0.437</td>
<td>(0.313)</td>
<td>0.445*</td>
<td>(0.300)</td>
</tr>
<tr>
<td>CEOPower</td>
<td>Entrenchment</td>
<td>+</td>
<td>-</td>
<td>0.179</td>
<td>(0.875)</td>
<td>1.416*</td>
<td>(0.811)</td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>+</td>
<td>-</td>
<td>0.002</td>
<td>(0.004)</td>
<td>-0.013*</td>
<td>(0.004)</td>
</tr>
<tr>
<td>U.S. Listed</td>
<td>Litigation</td>
<td>+</td>
<td>+</td>
<td>0.291</td>
<td>(0.264)</td>
<td>0.679*</td>
<td>(0.263)</td>
</tr>
<tr>
<td>Acquiror (%)</td>
<td>Litigation</td>
<td>+</td>
<td>+</td>
<td>-0.048*</td>
<td>(0.019)</td>
<td>0.004</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Divestor (%)</td>
<td>Litigation</td>
<td>+</td>
<td>+</td>
<td>0.071</td>
<td>(0.058)</td>
<td>0.004</td>
<td>(0.058)</td>
</tr>
<tr>
<td>Long term</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.890**</td>
<td>(0.083)</td>
<td>-0.130</td>
<td>(0.082)</td>
</tr>
</tbody>
</table>

| Firm-fixed effect | Included | Included |
| Time-fixed effect | Included | Included |
| Log Likelihood    | -1100    | -1054    |

The dependent variables are ordered policy limit and deductible. Ordered Probit regression using time- and firm-fixed effects. ** represents a coefficient that is significant at the 1% level, * significant at the 5% level and # at the 10% level. Some coefficients have been scaled for presentation purposes. Standard errors are in parentheses. N=696.
Because the shareholders' wealth and the reliance on debt as a source of capital appear to be positively linked to the demand for D&O insurance, I conclude that the evidence supports the view that D&O insurance is used by shareholders to protect their own wealth in case of managerial incompetence. Whereas the traditional literature on risk management predicts that larger corporations should purchase relatively less insurance because they face lower bankruptcy costs and are better able to spread risk over its different operations, I find that the more shareholders have at risk, the larger is the insurance protection.

Two other results are surprising by their absence in the sense that risk aversion and financial distress do not appear to find any support. In case of risk aversion, given that D&O insurance is generally said to attract better risk-averse directors unto the corporation's board, two possible reasons explain why risk aversion does not appear to play any role. First, it is possible that the variables I use are not good proxies for managerial risk aversion. A second reason is that D&O insurance is not used to reduce the risk-averse directors' risk exposure since better mechanisms are available to reduce it. One such mechanism is the use of limited liability plans for corporate boards. By adopting an LLP, a corporation tells potential directors that their personal wealth is not too much at risk in the event of a lawsuit brought unto them as representatives of the corporation. The LLPs typically state that outside directors cannot lose more than some relatively small amount in the event of a lawsuit. As a result, LLPs are better designed to alleviate the fears of outside directors than a D&O insurance. In case of financial distress, of all the measures I use, only the debt ratio had an impact that supported the hypothesis.

**References**


Becht, M., P. Bolton, A. Roell, 2003, “Corporate governance and control,” _In Constantinides, G., M. Harris, and R. Stulz (Eds.), Handbook of economics of finance, Elsevier, Amsterdam_


Chang, S.-C., and J. Yeh, 2011, “Show me the information: board independence and D&O insurance.” _Chinese University of Hong Kong working paper, 2011 ARIA meeting, San Diego_


Elfei, N. R., 2000, “Now you have it, now you don't: Directors' and Officers' insurance after a corporate bankruptcy.” Loyola Law Review 46, 585–617
Hanks, J. J., 1988, “Evaluating recent State legislation on Director and Officer liability limitation and indemnification,” Business Lawyer 43, 1207-1254
Kim, I., 2004, “The role of Directors' and Officers' insurance in financial reporting and class action securities litigation,” Mimeo, Duke University
Kremslehner, R., 2011, “Job market for Directors: the role of limited liability provisions and Directors' and Officers' insurance in optimal contracting.” WU – University of Vienna working paper, 2011 ARIA meeting, San Diego
LaCroix, K. M., 2012, “Canadian securities class action lawsuit filings hit record in 2011;” The D&O Diary
## Appendix

### Panel A: Deletion of lagged observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis</th>
<th>Limit Coefficient</th>
<th>St. Dev.</th>
<th>Deductible Coefficient</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash (1000)</td>
<td>Package</td>
<td>-</td>
<td>-7.252&quot;</td>
<td>(0.012)</td>
<td></td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress shareholder</td>
<td>+</td>
<td>0.077&quot;</td>
<td>(0.081)</td>
<td></td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress monitor shareholder</td>
<td>+</td>
<td>0.377&quot;</td>
<td>(0.003)</td>
<td></td>
</tr>
<tr>
<td>Blockholder (%)</td>
<td>Monitor</td>
<td>-</td>
<td>-0.671&quot;</td>
<td>(0.057)</td>
<td></td>
</tr>
<tr>
<td>Growth (%)</td>
<td>Signaling</td>
<td>+</td>
<td>(0.149)</td>
<td>0.314&quot;</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>+</td>
<td>(0.11)</td>
<td>0.181*</td>
<td>(0.105)</td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>-</td>
<td></td>
<td>-0.375&quot;</td>
<td>(0.129)</td>
</tr>
<tr>
<td>U.S.Listed</td>
<td>Litigation</td>
<td>+</td>
<td>0.231*</td>
<td>(0.104)</td>
<td></td>
</tr>
<tr>
<td>Long term</td>
<td>-</td>
<td>-0.186&quot;</td>
<td>(0.017)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Firm and time effects: Included

Goodness of fit: 0.954
Number of observations: 733

### Panel B: Inclusion of lagged observations

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis</th>
<th>Limit Coefficient</th>
<th>St. Dev.</th>
<th>Deductible Coefficient</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash (1000)</td>
<td>Package</td>
<td>-</td>
<td>-5.661&quot;</td>
<td>(2.237)</td>
<td></td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress shareholder</td>
<td>+</td>
<td>0.075&quot;</td>
<td>(0.028)</td>
<td></td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress monitor shareholder</td>
<td>+</td>
<td>0.439&quot;</td>
<td>(0.141)</td>
<td></td>
</tr>
<tr>
<td>Blockholder (%)</td>
<td>Monitor</td>
<td>-</td>
<td>-0.562&quot;</td>
<td>(0.167)</td>
<td></td>
</tr>
<tr>
<td>Growth (%)</td>
<td>Signaling</td>
<td>+</td>
<td></td>
<td>0.096&quot;</td>
<td>(0.105)</td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>+</td>
<td></td>
<td>0.146</td>
<td>(0.093)</td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>-</td>
<td></td>
<td>-0.33&quot;</td>
<td>(0.109)</td>
</tr>
<tr>
<td>U.S.Listed</td>
<td>Litigation</td>
<td>+</td>
<td>0.356&quot;</td>
<td>(0.100)</td>
<td></td>
</tr>
<tr>
<td>Long term</td>
<td>-</td>
<td>-0.170&quot;</td>
<td>(0.018)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Firm and time effects: Included

Goodness of fit: 0.936
Number of observations: 884

### Table A1: The determinants of policy limit and deductible amounts restricted model OLS regression with firm and year fixed effects

The dependent variables are the log of the policy limit and the log of the deductible. OLS regression for limit and deductible using time and firm fixed effects. Number of determinants restricted. The hypothesis or the hypotheses that each independent variable tests is presented in the second column. Also, the second column gives the predicted sign of each coefficient on the dependent variables. The ** represents a coefficient that is significant at the 1% level, * significant at the 5% level and # at the 10% level. Some coefficients have been scaled for presentation purposes. Standard errors are in parentheses.
Table A2: The determinants of policy limit and deductible amounts OLS regressions using firm and year fixed effects (deductible chosen after limit)

The dependent variable is the log of the deductible. Sequential model where the insured knows the policy limit when he chooses the deductible. OLS regression using time and firm fixed effects. Only the deductible regression is presented since the limit regression is the same as in Table 4. The ** represents a coefficient that is significant at the 1% level, * significant at the 5% level and # at the 10% level. Some coefficients have been scaled for presentation purposes. Standard errors are in parentheses. N=696.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Hypothesis</th>
<th>Predicted sign</th>
<th>Coefficient</th>
<th>Deductible</th>
<th>St. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash (1000)</td>
<td>Package</td>
<td>+</td>
<td>-0.154</td>
<td>(1.158)</td>
<td></td>
</tr>
<tr>
<td>Option</td>
<td>Package</td>
<td>+</td>
<td>-0.028</td>
<td>(0.025)</td>
<td></td>
</tr>
<tr>
<td>Stock (1000)</td>
<td>Package</td>
<td>+</td>
<td>0.037</td>
<td>(1.158)</td>
<td></td>
</tr>
<tr>
<td>Independent (%)</td>
<td>Aversion</td>
<td>-</td>
<td>1.204</td>
<td>(2.324)</td>
<td></td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress</td>
<td>+</td>
<td>0.008</td>
<td>(0.063)</td>
<td></td>
</tr>
<tr>
<td>Bankruptcy (%)</td>
<td>Distress</td>
<td>-</td>
<td>-0.728</td>
<td>(0.467)</td>
<td></td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress</td>
<td>+</td>
<td>0.333</td>
<td>(0.288)</td>
<td></td>
</tr>
<tr>
<td>RoA</td>
<td>Distress</td>
<td>+</td>
<td>-0.047</td>
<td>(0.252)</td>
<td></td>
</tr>
<tr>
<td>Blockholder (%)</td>
<td>Monitor</td>
<td>+</td>
<td>-0.176</td>
<td>(0.362)</td>
<td></td>
</tr>
<tr>
<td>Growth (%)</td>
<td>Signaling</td>
<td>+</td>
<td>0.396**</td>
<td>(0.138)</td>
<td></td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>+</td>
<td>0.234#</td>
<td>(0.125)</td>
<td></td>
</tr>
<tr>
<td>CEO Power</td>
<td>Entrenchment</td>
<td>-</td>
<td>0.344</td>
<td>(0.346)</td>
<td></td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>-</td>
<td>-0.456**</td>
<td>(0.161)</td>
<td></td>
</tr>
<tr>
<td>U.S. Listed</td>
<td>Litigation</td>
<td>+</td>
<td>0.214*</td>
<td>(0.109)</td>
<td></td>
</tr>
<tr>
<td>Acquiror (%)</td>
<td>Litigation</td>
<td>+</td>
<td>0.428</td>
<td>(0.991)</td>
<td></td>
</tr>
<tr>
<td>Divestor (%)</td>
<td>Litigation</td>
<td>+</td>
<td>-0.335</td>
<td>(2.443)</td>
<td></td>
</tr>
<tr>
<td>Long term</td>
<td>Entrenchment</td>
<td>-</td>
<td>-0.022</td>
<td>(0.036)</td>
<td></td>
</tr>
<tr>
<td>Limit</td>
<td></td>
<td></td>
<td>0.143</td>
<td>(0.089)</td>
<td></td>
</tr>
<tr>
<td>Firm-fixed Effect</td>
<td></td>
<td></td>
<td>Included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-fixed Effect</td>
<td></td>
<td></td>
<td>Included</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goodness of fit</td>
<td></td>
<td></td>
<td>0.928</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Directors’ and officers’ insurance and shareholder protection

#### Table A3: The determinants of policy limit and deductible amounts

Instrumental variables, two-stage regression with firm and year fixed effects

The dependent variables are the log of the policy limit and the log of the deductible. The number of determinants restricted. Two-stage regression using instrumental variables for limit and deductible, and time- and firm-fixed effects; reduced equation not shown. Panel A presents the regression excluding the lagged variables whereas Panel B presents the regression including the lagged variable. The ** represents a coefficient that is significant at the 1% level, * significant at the 5% level and # at the 10% level. Some coefficients have been scaled for presentation purposes. Standard errors are in parentheses.

<table>
<thead>
<tr>
<th>Panel A: Deletion of lagged variables; N=700</th>
<th>Pred. sign</th>
<th>Policy limit</th>
<th>Deductible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash (1000)</td>
<td>Package</td>
<td>-8.130* (2.446)</td>
<td></td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress shareholder</td>
<td>0.081* (0.028)</td>
<td></td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress monitor shareholder</td>
<td>0.474** (0.028)</td>
<td></td>
</tr>
<tr>
<td>Blockholder (%)</td>
<td>Monitor</td>
<td>-</td>
<td>0.325** (0.115)</td>
</tr>
<tr>
<td>Growth (%)</td>
<td>Signaling</td>
<td>+</td>
<td>0.176* (0.105)</td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>+</td>
<td>-0.380** (0.129)</td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>-</td>
<td>0.196* (0.105)</td>
</tr>
<tr>
<td>U.S.Listed</td>
<td>Litigation</td>
<td>+</td>
<td>0.169** (0.018)</td>
</tr>
<tr>
<td>Long term</td>
<td>-</td>
<td>-0.169** (0.131)</td>
<td></td>
</tr>
<tr>
<td>Predicted (Limit)</td>
<td></td>
<td>0.221* (0.131)</td>
<td></td>
</tr>
<tr>
<td>Firm and time effects</td>
<td></td>
<td>Included</td>
<td>Included</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td></td>
<td>0.957</td>
<td>0.926</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Inclusion of lagged variables; N=839</th>
<th>Pred. sign</th>
<th>Policy limit</th>
<th>Deductible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cash (1000)</td>
<td>Package</td>
<td>-8.808** (2.211)</td>
<td></td>
</tr>
<tr>
<td>LnMVE</td>
<td>Distress shareholder</td>
<td>0.077** (0.028)</td>
<td></td>
</tr>
<tr>
<td>DebtRatio</td>
<td>Distress monitor shareholder</td>
<td>0.317** (0.106)</td>
<td></td>
</tr>
<tr>
<td>Blockholder (%)</td>
<td>Monitor</td>
<td>-</td>
<td>0.150 (0.093)</td>
</tr>
<tr>
<td>Growth (%)</td>
<td>Signaling</td>
<td>+</td>
<td>-0.350** (0.109)</td>
</tr>
<tr>
<td>Volatility</td>
<td>Signaling</td>
<td>+</td>
<td>0.323** (0.101)</td>
</tr>
<tr>
<td>LnWealth (1000)</td>
<td>Entrenchment</td>
<td>-</td>
<td>0.161** (0.018)</td>
</tr>
<tr>
<td>U.S.Listed</td>
<td>Litigation</td>
<td>+</td>
<td>0.043 (0.118)</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td></td>
<td>0.943</td>
<td>0.912</td>
</tr>
</tbody>
</table>
A long-term approach to Italian banks’ profitability: paradise lost?

Giuseppe Lusignani
Professor of Economics and Financial Intermediaries, University of Bologna

Marco Onado
Senior Professor, Department of Finance, Bocconi University

Abstract
This paper examines the long-term profitability of the Italian banking system (1965-2012), which has three interesting features: it has always been oriented toward traditional intermediation; it underwent a significant process of privatization, consolidation and modernization; and it proved to be very robust when the financial crisis took place. However at present, both RoAs and RoEs have reached the low levels of previous troughs. In a sense, the Italian banking system seems to be back to square one. Our analysis shows that the most important driver of profitability has been the decline of the margins since the late 1980s and that a reversal of this trend does not seem likely in the foreseeable future. Consequently, future rebounds in profitability can only come from significant actions of rationalization, which do not seem to have been delivered during previous merger waves.

1 This article enriches, modifies and updates a previous study, which was part of the overall research report “Euro shocks and the Italian banking system” organized by the Rosselli Foundation and published yearly by Bancaria (Italian Banking Association). We wish to thank Prometeia for providing the data; Sara Emiliani, Teresa Sardena and Emanuele De Meo for excellent research assistance, and Andrea Enria, Andrea Resti, Andrea Sironi, Angelo Tantazzi and Lea Zichino for helpful comments.
1. Foreword
The financial crisis has brought the long-term profitability of the banking industry to the fore and raised many questions. These include, how have the structural changes of the financial markets over the past 50 years (liberalization, globalization, innovation, and for the euro area countries, monetary union) changed banks’ profitability? Was there a substantial difference in terms of sustainability between the profitability of the traditional intermediation business (utility banking in Lord Turner’s definition [Turner Review (2009)]) and the investment banking business? Were the profits recorded in the run up to the crisis only a mirage [Haldane (2009)]? And from a forward-looking perspective, what are the implications for future levels of profitability?

The crisis has shown the flaws of many business models that banks adopted during the long phase of credit expansion and innovation. Victims can be found both in banks that relied too much on risky trading activities, as well as those whose retail banking model was compatible only with the rosy scenario of eternal growth and limitless availability of liquidity.

The Italian banking system is an interesting case in point: historically, it has focused on the traditional intermediation between households, companies and — on a smaller scale in the last decades — the public sector. Even the international expansion of its two leading banks has been mainly based on acquisitions of retail banks, particularly in East European countries. In the 1990s, it also underwent the passage from public to private ownership along with a significant consolidation.

Italian banks proved to be among the most sound at the onset of the crisis: only a few of them had to ask for state-guaranteed bonds and for small amounts (both in absolute terms and relative to their size). But as the crisis continues, with the end, at least for the periphery of Europe, still some way away, some certainties have begun to falter. It is, therefore, interesting to look at the long-term evolution of the profitability of the Italian banking system as a typical example of the effects of the profound changes of the last decades on the traditional intermediation model and also as a case in point of the challenges that European banks must learn to manage in the foreseeable future.

An implicit hypothesis of a study on aggregate banking profitability is that in each country, there are idiosyncratic market drivers of the general trend of both revenues and costs that are among the major constraints impacting individual bank’s strategic choices. As a consequence, the average trend of profitability is a major determinant, perhaps the most important one, of margins for any bank.

Our analysis is based on a reliable long-term database of the main indicators of profitability for the entire Italian banking
system, thanks to the data provided by the Bank of Italy since the mid 1960s, a wide body of previous research (Passacantando (1983), Onado (1986), Angeloni et al. (1999), Albertazzi and Gambacorta (2009)) and the statistical database used by Prometeia for its analysis and forecasts of banks’ profitability (Prometeia, Analisi dei bilanci bancari and Previsione dei bilanci bancari).

The paper is organized as follows. The next section reviews the stylized facts and data on the evolution of the Italian banking system, providing a snapshot of today’s characteristics. Section 3 examines the long-term trend of the main margins, while the following section focuses on the two main indicators of gross profitability: net interest income and net income. Section 5 deals with the cost structure and operating efficiency. Section 6 focuses on the items below gross operating profit and, in particular, on loan loss provisions and on the two main indicators of the overall profitability, namely return on assets (RoA) and return on equity (RoE). Section 7 looks at some indicators of profitability for the major groups of banks by asset size. Finally, conclusions and implications for future prospects for the Italian banking system are presented in Section 8.

2. The Italian banking system: a few stylized facts

It is worth remembering the main characteristics and the long-term trend of the Italian economy and its financial system.

A. An economy with many contradictions

Italy’s growth rates were amazingly high in the first 15 years after World War II (the so-called Italian miracle) and then declined, particularly in the past two decades (Figure 1). The inflation rate (measured by the consumer price index) shows a reverse V-shaped pattern: prices soared after the first oil shock and remained high (and above the average for advanced countries) throughout the 1980s. It also explains the gap with real growth. The following decades were dominated by the financial crisis of 1992, which led to the heavy fiscal adjustment of the following years, and that opened the path to the admission of Italy to the monetary union in the first phase (1998).

After the first oil shock, Italy experienced an inflation wave, which was longer and more severe than other industrialized countries. The country’s growth model based on rising prices and continuing devaluation, public budget deficit and rising

Government debt was interrupted only at the beginning of the 1990s (in the aftermath of a major currency crisis) when Italy strived to meet the conditions to enter the monetary union, to which it was eventually admitted in 1998 (Figure 2). In the run-up to the admission, interest rates converged rapidly toward other countries’ levels.
The Italian production structure is heavily export-oriented; hence, the long inflation wave (which was significantly higher than in other countries) led to successive devaluations of the lira (Figure 3). Each devaluation restored the competitiveness of the Italian export industry and led to significant surpluses of the current accounts balance. In the recent past, the trade balance turned negative again, as the monetary union did not allow the country to offset its decreasing competitiveness with a devaluation. The effect is even more pronounced in the current account balances, which mirror the decreasing competitiveness of the service sectors, namely tourism, and the growing flow of interest payments on Italian debt held abroad.

The increasing current account deficit also points to an increasing demand for foreign capital and, therefore, for example, an increasing intermediation of international financial flows. As bank loans increased much faster than retail funding, as will be discussed below, Italian banks were forced to tap international (bond and interbank) markets, which as a result became an important source of the Italian demand for foreign capital.

Another feature of the Italian economy has been the high saving ratio, particularly of the household sector (Figure 4). However, the households’ financial surplus decreased steadily over the last decades, following a pattern similar to other countries. Nevertheless, Italian households’ financial wealth, at the end of the period, is higher, as a percentage of GDP, than that of the other advanced countries [De Bonis (2007), Bartiloro et al. (2008)].

B. A significant growth in the size of the banking sector, but lower than in most other European countries

Figure 5 shows a significant increase in the size of the banking system (measured by total assets) since the early 1950s. In nominal terms, total assets, which amounted to less than €500 billion at the beginning of the 1980s, trebled in a decade and doubled during the following two, in parallel with the creation of the monetary union and the international credit boom that led to the financial crisis. Figure 5 also shows the level of nominal GDP and the ratio of banks’ total assets to GDP. This measure of banks’ size, normalized by GDP, demonstrates three distinct phases: the strong recovery of the first two decades (reaching 1.5 in the mid 1970s); a slowdown in the 1970s, 1980s and early 1990s (which means that the asset growth was driven by inflation); and an acceleration in the next two decades. However, over the last few years, this has hovered around 2.5, which is much lower than major Euro area banking systems (Figure 6).

The Italian production structure is heavily export-oriented; hence, the long inflation wave (which was significantly higher than in other countries) led to successive devaluations of the lira (Figure 3). Each devaluation restored the competitiveness of the Italian export industry and led to significant surpluses of the current accounts balance. In the recent past, the trade balance turned negative again, as the monetary union did not allow the country to offset its decreasing competitiveness with a devaluation. The effect is even more pronounced in the current account balances, which mirror the decreasing competitiveness of the service sectors, namely tourism, and the growing flow of interest payments on Italian debt held abroad.

The increasing current account deficit also points to an increasing demand for foreign capital and, therefore, for example, an increasing intermediation of international financial flows. As bank loans increased much faster than retail funding, as will be discussed below, Italian banks were forced to tap international (bond and interbank) markets, which as a result became an important source of the Italian demand for foreign capital.

Another feature of the Italian economy has been the high saving ratio, particularly of the household sector (Figure 4). However, the households’ financial surplus decreased steadily over the last decades, following a pattern similar to other countries. Nevertheless, Italian households’ financial wealth, at the end of the period, is higher, as a percentage of GDP, than that of the other advanced countries [De Bonis (2007), Bartiloro et al. (2008)].
In summary, the growth of Italian banks can be divided in three periods: (i) the strong recovery during the “Italian miracle”; (ii) the big inflationary wave (iii) the years of the monetary union. It is worth stressing that in real terms the last decade has seen a remarkable acceleration of the ratio of total assets to GDP, but this phase of “financial deepening” did not coincide with a period of strong economic growth. Finally, in those years, the two major banking institutions made important acquisitions abroad, which increased their total assets (the foreign component of total assets is not included in the data of Figure 5).

C. An extraordinary effort of deregulation and privatization, prompted by the Bank of Italy since the 1980s
At the beginning of the 1980s, the Italian banking system had three main characteristics: (i) it was publicly owned: almost 80% of banks were either government-owned or public bodies (they included the savings banks, Banco di Napoli, Banco di Sicilia, Istituto San Paolo di Torino, Monte dei Paschi di Siena, and all of the institutions specialized in long-term lending); (ii) it was heavily regulated by the Banking Act (1936), which created many segmentations (between short- and long-term lending; between bank categories), posed strict limits on branching and gave strong administrative powers to the Bank of Italy and (iii) monetary policy relied on administrative controls instead of open market operations.

In the 1970s, administrative constraints (vincoli di portafoglio) made it mandatory for the banks to invest a significant part of their deposits in bonds issued by “istituti di credito speciale” (public bodies specialized in long-term credit), while loans were subject to severe credit ceilings (massimale). During this period, Italian banks were less enterprises responding to market incentives than the “chain of transmission” of a monetary policy aimed at facilitating the smooth refinancing of the public deficit. Moreover, controls on foreign flows of capital and holdings by residents of foreign-denominated assets were enforced for most of those years.

The picture changed dramatically during the 1980s thanks to the debate on the public banks prompted by the Bank of Italy, the gradual lifting of administrative controls and successive amendments to the Banking Act. As for banks’ ownership, the breakthrough was the reform approved at the beginning of the 1990s (the so-called Amato Law) under which public banks’ assets and liabilities were transferred to newly created joint-stock companies (therefore, of private nature) and the ownership of the capital was attributed to a foundation (the so-called Fondazioni bancarie). Then, the foundations were pushed to sell a significant part of their holdings, thus favoring the enlargement of the shareholder base. Meanwhile, as a consequence of the implementation of European directives, the banking legislation and regulation were gradually changed, abolishing the many forms of segmentation of the previous Banking Act.

D. A traditional business model, based on deposit-loan intermediation
The importance of the traditional intermediation is shown in Figure 7. The share of loans to total assets follows two completely different trends. First, during a long phase of continuous decline, banks were supporting the bond market through direct acquisitions of government bonds and securities issued by credit institutions, while loans were subject to severe credit ceilings. At the end of the 1980s, administrative controls were lifted and since then, the importance of loans in the overall portfolio has continuously risen. At present, Italy has one of the highest ratios of loans to total assets among the European banking systems, lower only than Spain and the U.K. among the large countries.
A long-term approach to Italian banks’ profitability: paradise lost?

At the same time, a wide body of literature shows that customer relationship has been significantly strengthened as a consequence of the process of consolidation [Bonaccorsi di Patti and Gobbi (2001, 2007)].

In recent years, loans have grown at a much faster pace than the traditional retail sources of funding (particularly deposits). As Figure 7 shows, the ratio of loans to deposits, which was about 100% at the beginning of the 1990s, jumped to 177 in the run up to the financial crisis. It has fallen since the crisis to 143. The funding gap of the Italian banking system is therefore significant, as in most European countries. However, Italian banks sell a large quantity of bonds to their retail customers, at prices that are much cheaper than those offered to institutional investors. Sadly, available data do not allow for separating the retail and wholesale components of bond issues.

3 According to an estimate by Goldman Sachs (2011), by mid 2011 retail bonds represented 40% of total bonds issued by the top 8 Italian banks (the percentage being lower for the two largest bank).

The gap between the ratio of loans to funding sources (loans to total amount of deposits and bonds) has widened in recent years, reflecting a heavy reliance of Italian banks on issuance of securities. At the peak (2012), bonds were equal to almost 55% of the total loan portfolio (or 44% of total funding sources). The retail funding gap proved to be “one of the most important sources of banking sector vulnerability throughout the financial crisis” [ECB (2011)].

E. A significant strengthening of the capital base

Capitalization, as measured by the ratio of capital to total assets, is one of the key features of a banking system. In countries where long-term series are available, it has been shown [Alessandri and Haldane (2009)] that capital declined steadily since the 19th century, reaching its lowest levels by the end of WWII. Since then it has hovered around 5% (Figure 9). In other words, the last three decades, characterized by the implementation of Basel rules on capital adequacy, could only align the growth of capital with the extraordinary growth of total assets. Comparable data are not available for Italy for such a long period. Istat has provided a series since Italian unification (1861) for capital, loans, securities and real estate. Using the sum of the last three data as a proxy for total assets, the capitalization shows a sharp and steady decline from 1861 to 1950 and then a strong rebound [Onado (2012)]. However, this estimate suffers from the heroic assumption that the other components of bank balance sheet remain constant and is therefore not homogenous with the series used for the U.S. and the U.K. Moreover, it must be taken into account that Italy had a long tradition of publicly owned banks, where the equity protection was less significant. Consequently, we have included in the graph the capitalization ratio for Italy since the early 1950s.
During the 1950s and 1960s, the Italian banking system was much less capitalized than its Anglo-Saxon peers. Since then, however, it underwent a long phase of playing catch-up. By the early 1990s, levels of capital within the Italian banking sector were more or less in line with those of the U.K. and U.S. Even after this period, Italian banks’ capital continued to grow faster than total assets, so much so that in recent years Italian banks had significant higher levels of capital than their American or British peers.

F. A significant concentration process
Thanks also to the privatizations, the last 20 years have seen a strong wave of mergers and acquisitions. The number of banks declined significantly, from more than 1,100 in 1990 to slightly above 700 in 2012, albeit at a slower pace than in Germany and France (Figure 10). Nevertheless, concentration measures (the Herfindahl index and the market share of the five largest institutions) are significantly lower than in other European countries, excluding Germany (Figure 10). The concentration has pushed leading Italian banks up the international banking league tables, with Unicredit and IntesaSanpaolo being ranked 23rd and 28th, respectively, in 2011.4

3. Profitability from a long-term perspective: a bird’s-eye view
In this section, we look at the long-term trends of the major banking margins, with a deeper analysis in the following sections. We adopt the traditional definition (OECD 1979–2010) of the main components of banks’ profit and loss statements (see Box 1 for a brief description).5

Figure 12 presents the long-term trend for the four main margins since 1965. The main points to be stressed are:

- Net interest margin follows two different patterns: one of high and fluctuating margins (1965–90), and one of continuous decline in the last two decades. For a system strongly orientated toward traditional intermediation, this means a

---

4 According to Top Banks by assets 2011, Unicredit is in 23rd position and Intesa Sanpaolo in 28th. (http://www.relbanks.com/worlds-top-banks/assets-2011); in Bankscope global ranking, Unicredit is ranked 23rd and Intesa Sanpaolo 32nd.

5 As in the prevailing literature, we use total assets as the variable of scale to measure bank output. This measure is far from flawless: (i) it is a spurious mix of book and mark-to-market values, depending on accounting standards; (ii) it is unconsolidated, therefore it can be inflated by assets that would be cancelled if we would add up individual figures as in a consolidation process; (iii) its composition varies over time, therefore, the underlying risk is not constant. Having said that, it must be recognized that it is very difficult to find an alternative to total assets as the denominator of profitability ratios. For a system heavily orientated toward traditional intermediation, one can use items strictly connected to this business, namely loans and deposits (for example, the sum of the two). However, this indicator seems to follow rather strictly the pattern of total assets; the correlation ratio between the two series in our sample period is 0.997.
structural and long-term deterioration of the profitability of the basic business model.

- The trend for net income is rather similar. In particular, the graph shows the steady decline during the last decade. In other words, non-interest income (the distance between the two curves) increased significantly, but not enough to compensate for the long-term decline in interest margins.

- Gross operating profit (before provisions) – a measure of gross profitability that is not influenced by the cycle of credit losses – fluctuates over the period around a long-term average of about 1.5 and declines markedly since the beginning of the new millennium.

- Profit before taxes (PBT) – a measure of gross RoA, which is not influenced by the rate of taxation – experienced two ascending trends (1974-1991 and 1994-2007), interrupted by two major crises. In each cycle, the decline brings the overall profitability to the starting level. Moreover, both in 2011 and 2012 the bottom-line result of the entire banking system has been a loss. The first positive trend was caused by the inflation shocks of the 1970s and 1980s, while the second was brought about by the monetary union and the restructuring of the banking system. Both were interrupted by a recession linked to monetary causes (the currency crisis of the early 1990s and the European crisis with a sharp increase in Italian interest rates). The net result is that at the end of the period, the gross profitability of Italian banks seems to go back to square one and has recorded for the first time in recent years two consecutive years of net losses.

4. The components of gross profitability: NIM and NI

4.1 The overall trend of NIM

The Italian banking system has seen a long and positive interest rate cycle during the 1960s and 1970s, i.e., in the years of high inflation and high interest rates, which determined a real “intercept jump” of the level of interest margins [Onado (1986)]. The highest level was achieved in 1975 (3.9%), with minor fluctuations during the 1980s; in 1990 NIM still stood at 3.7%. Since then, the decline has been continuous. From 1991 to 2011, with the exception of 2006, NIM fell every year. In one decade, it lost more than 160 basis points, reaching 1.9% in 2000; another 90 basis points were lost in the following decade, dropping to 96 basis points in 2012, the lowest level for the entire period. The final margin is equal to only 30% of the peak reached 30 years earlier. Not surprisingly, the era of declining inflation and declining interest rates (thanks to the monetary union in the last two decades) has profoundly affected the basic profitability of the Italian banking system, leading to a constant decline of interest margins.
Moreover, no significant differences can be seen before and after the financial crisis that began in 2007. In particular, Italian banks did not benefit from the sharp and continuing decrease of interest rates coupled with a deepening of the yield curve. We will elaborate on this point in next section.

4.2 The determinants of interest margins: NIM, mark-up and mark-down on market interest rates

NIM is obviously strictly dependent on lending rates \( r_L \) and deposit rates \( r_D \) and on their difference, the spread. As our data cover the entire period, it is interesting to analyze the traditional relationship that expresses the spread as the sum of a markup \( (r_L - r_M) \) and a markdown \( (r_M - r_D) \) to a short-term market rate \( r_M \).

As Figure 13 shows, the spread increased significantly in the period of high inflation and high interest rates, reaching its peak at 15 percent in 1981. The decline was then almost uninterrupted; in 1990 it was still above 10%, but by the turn of the millennium it had reached about 5%. In 2012, the latest available figure, the spread was 4.7%, which is a third higher than the record level reached 30 years ago. In other words, the price component of interest margin has been dramatically reduced.

The decline of the spread was accompanied by a significant change in the balance between the mark-up and the mark-down. Until the late 1990s, the latter made a significant contribution to the short-term spread, accounting for about 60% in the 1970s to less than 50% in the 1980s. In the last 20 years, the deposits markdown contribution to the short-term spread kept declining: before the crisis, it was less than 40% of the total and in the last two years it ranged from 10% to 15%.\(^6\)

The reason must be found in the low level of market interest rates (which puts a floor to the markdown) and in the more competitive conditions of banking markets in the last decades. The net effect is that in the present market environment, borrowers (both households and companies) contribute almost entirely to this component of the profitability of the Italian banking system.

\[^6\] As figure 13 shows, when we account for the cost of bond borrowing, the contribution of markdown (calculated as a difference of short term money market interest rates and the weighted average of interest rates on deposits and bonds) to the spread is even lower. Moreover, in the last two years this contribution has become negative.

4.3 The determinants of interest margins: macro variables and market rates

Given the importance of interest margins in determining the long-term profitability of Italian banks, it is interesting to further investigate the determinants of NIM.

With regards to the impact of interest rates, Albertazzi and Gambacorta (2009) find that the coefficient for the money
market rate is close to zero. This result corroborates those of Demirgüç-Kunt and Huizinga (1999) and Casolaro and Gambacorta (2005). On the other hand, the coefficient for the long-term interest rate turns out to be positive and significant: if the long-term interest rate rises by 1 percentage point, the net interest margin is expected to increase by more than 1% in the first year and by almost 4% afterward. The different impact of short- and long-term interest rates on net interest income is due to the maturity transformation typical of banking.

More recently, the ECB has estimated the elasticity of the interest income of banks in the euro area to short-term market rates for the period from January 2003 to December 2008 (ECB (2009, box 13)). The results are particularly interesting and can be summarized as follows: (i) the impact of short-term market rates is stronger on loan rates than on deposit rates; (ii) the standard deviation of the estimated coefficient is however quite high (loan rates’ elasticity varies between 0.05 and 0.60, while deposit rates’ between 0.06 and 0.42); (iii) there is a marked difference between countries whose banks offer predominantly long-term fixed loans (Belgium, Germany and France) and those, such as Italy, with predominantly floating rates and short-term lending (in the first group of countries, the multiplier coefficient is higher for deposit rates than for lending rates) and (iv) the overall effect on interest margins is strongly related to the funding gap, which is smaller in countries where fixed rates prevail; hence the net effect turns out to be slightly positive.

Empirical studies conducted by the Bank for International Settlements also examine the possible effects of changes in market interest rates – both the level and the slope of the yield curve – on banks’ profitability. In a 2002 empirical exercise (BIS (2002), the analysis, based on aggregate data of net interest margin, was focused on the long-run behavior of commercial banks in 10 industrial countries, while in a more recent study (BIS (2012)) the focus has been on the short-run effects of the very aggressive and protracted monetary accommodation by central banks subsequent to the recent crisis on the profitability of a specific sample of internationally active banks.

Even if the conventional view is that changes in the level of interest rate and in the slope of the yield curve should have significant and positive effects on banks net interest income, the two studies provide mixed results for the sign of the slope: the correlation between bank’s net interest margins and the slope of the yield curve depends on how rapidly and how symmetrically the banks are able to re-price their liabilities and their assets, when the interest rates change. However, in the most recent empirical work (BIS (2012)), the coefficient estimated by running a cross-section regression over the period 2008-10, shows a positive correlation between the net interest margin and both the level of short-term interest rates (0.258) and the slope of the yield curve. This means that in a scenario of low market rates, banks can find profitability only through maturity transformation.

With regards to the macro variables, previous studies of the Italian banking system have found a significant relationship between interest income and the economic cycles [Albertazzi and Gambacorta (2006)]. In particular, GDP turns out to be significant and positive. When GDP increases by 1%, net interest income increases by 0.6% in the first year. The effect is even more pronounced in the long run, reaching 1.8%, probably due to the presence of long-term contracts. The economic interpretation is simple: an improvement in economic conditions increases lending demand by households and firms [Friedman and Kuttner (1993), Calza et al. (2003)] and improves the financial conditions of borrowers, with positive effects on the profitability of the traditional financial intermediation activities.

Our goal is slightly different from previous research, because we are not only interested in very short-term movements (as in the ECB and BIS studies) or cyclical behavior, we are also interested in long-term and structural effects. Our analysis is performed in two steps. First, we investigate the simple relationship between NIM and short-term market rates as well as the slope of the yield curve, measured by the difference between the 10-year and the 3-month money market rate.

---

7 The 2002 study ran 10 different regressions over the period 1980–2001 between the net interest rate margin and the market interest rate of 10 different countries, and could not find strong evidence to support the proposition that the slope has significant and positive effect on banks’ income. Only in the case of U.S. does the slope of the yield curve significantly impact, and with a positive sign, the net interest margin. The slope of the yield curve coefficient is statistically significant, but negative, in Germany, Sweden, Norway and Switzerland. In the case of Italy and Canada, no significance was found in the relationship, though the sign was negative (BIS (2002)). The 2012 study (BIS (2012)), however, found that the coefficient estimated by running cross-section regressions over the period 2008-10 demonstrated a positive correlation between net interest margin and both the level of short-term interest rates (0.258) and the slope of the yield curve (0.641 the coefficient to slope of the 0-2 years yield curve, and 0.820 the coefficient to slope of the of the 2-10 years yield curve).
We estimate the relationship looking at the NIM elasticities, i.e., the percentage change of the NIM in relation to a given change in market rates according to the following specifications:

\[ \text{NIM}_t = \alpha_t + \beta_t \text{MR}_t + \eta_t \]
\[ \text{NIM}_t = \delta_t + \gamma_t (\text{LR}_t - \text{MR}_t) + \epsilon_t \]

where MR$_t$ is the short-term money market rate at time $t$ and LR$_t$ is the long-term interest rate at time $t$; \( \alpha_t \) and \( \delta_t \) are constant – which are estimated at each time $t$ – and \( \eta_t \) and \( \epsilon_t \) are random errors, which are normally distributed with zero mean and standard deviation equal to 1.

The two relationships are estimated separately in order to avoid collinearity problems. Figure 14 shows the results of the Kalman filter estimation on a recursive data sample of annual data (from 1970 to 2012). This technique gives us an immediate snapshot of the changes in the NIM elasticity.

Our results corroborate previous findings on the positive relation between NIM and the level of short-term interest rates; this relationship is increasing from the early 1980s showing a stronger sensitivity of NIM to money market conditions. The value of alpha captures all other determinants of NIM, including the competitive conditions of the banking markets. The increase from the early 1970s through the late 1980s and the sharp decline thereafter can be considered as the effect of a permanent increase of competition favored by the structural changes that began in the early 1980s. As far as the yield curve slope is concerned, the estimates show a negative relationship, suggesting that liabilities respond more quickly than assets to market rates. Stated differently, mainly because of the demands of depositors, Italian banks seem to perform a negative asset-liability transformation. Summing up, our results seem to point to a very unfavorable environment for Italian banks in the past decade, as both the level and the slope of the interest rate curve have had a negative impact on NIM. Our results, therefore, depict a more negative picture than the ECB analysis discussed earlier.

Since this finding could be the result of either a difference in the data or a difference in the model, we ran an estimate similar to that described in the ECB (2009), i.e., an error-correction regression of the change in the average interest rate paid on outstanding loans and deposits respectively on changes in the three-month short rate.

This second exercise allows us to further investigate the elasticity of the main components of the interest margin (rates on loans, deposits and bonds) to market interest rates and, thanks to
the error-correction model, to separate the long-term from the short-term effects of a given change in market interest rates. In fact, in an error-correction model the short-term dynamics of the variables are influenced by their distance from the equilibrium values. In particular, we estimate the following equation:

$$\Delta BR_t = a + \gamma (BR_{t-1} + \theta MR_{t-1}) + \sum_{i>0} \beta_i \Delta MR_{t-i},$$

where $BR$ is alternatively the rate on loans or deposits or bonds, $MR$ is the short-term money market rate and $\Delta$ indicates the first difference. The above equation has been estimated for a sample of quarterly data from 1970 to 2012 as well as for different sub-periods to examine whether the elasticities change over the sample period. In Table 1, we show our estimates of both short- and long-term elasticities over the different periods.

The estimates of short-term elasticities are very similar to those obtained by the ECB, when the model is estimated over the same sample period. This makes us more confident about the model and allows us to concentrate on the long run elasticities. The results on long-term elasticities show a striking difference between the first period (1970–1984), where rates rise, and the second period (1985–2012), where rates fall. In the former, loan rates reacted faster than deposit rates (1.28 versus 0.72), thereby giving a positive contribution to interest margins. In the latter, with declining interest rates, even if loan and deposit rate elasticities are lower than the previous period (1.09 and 0.53, respectively), the contribution to interest margin was negative. In particular, in the period after the introduction of the euro (1998–2012) loan rates and deposit rates elasticities were very close, with a negligible effect to interest margins.

This result can be attributed on the one hand to the low level of market rates (it is progressively more difficult for banks to cut rates when they are approaching the zero bound) and, on the other hand, to the more competitive conditions, particularly in the deposit markets. As a matter of fact, in the more recent decades, the competition has been very tough, both from nonbank sources (the government bond market, institutional investors, etc.) and within the banking system, as loans grew much faster than deposits, creating the conditions for price competition in the deposit market.

Generally, the variability of interest rates is considered per se to be beneficial to bank interest margins [Albertazzi and Gambacorta (2006)]. The main reason is that the speed at which changes in market interest rates are incorporated into the bank interest margin differs over time and, usually, the adjustment takes place fairly gradually. Consequently, the long run elasticity (i.e., the equilibrium elasticity in a co-integrated model) is more relevant than the short one, and the long-term trend is more important than the short-term movements. Our results show that this was not true for the Italian banking system in recent years. The long-term forces driving the decline of margins seem to have been stronger than the short-term movements in the level and slope of the interest rates curve.

### 4.4 Net Income

Figure 12 shows that net income (NI) also declined steadily over the last 30 years, albeit at a slower pace than NIM. NI stood at 1.9% in 2012, 39% of the peak level reached in 1981 and 1986. In the last decade, non-interest revenues declined at the same pace as NIM (in 2011 they were both 54% lower than in 2001). As a matter of fact, the decline of NIM was far from unexpected (even though not for such an extended period and so uninterrupted). Italian banks were only too well aware that in the long run they should change their traditional business model, finding new sources of revenue such as trading and fees for non-traditional services. Diversification seemed sustained by the increased propensity of households to invest in financial assets other than government bonds, and by the steady development of financial markets, which widened the range of services to be offered both to companies and households [Enria et al. (1999, Albertazzi and Gambacorta (2006)].

Figure 15 shows the trend of non-interest income (NII, the difference between NI and NIM) both in absolute value and as a percentage of total assets. The nominal value increases steadily over the period: during the 1990s, the level trebled and after a pause due to the bursting of the stock market bubble, jumped again to the record level of more than €45b. The ratio of NII to total assets, which stood at 1.2% at the beginning of the 1980s, declined to 0.6 in 1991 and then increased very fast, jumping to 1.6 at the turn of the millennium, then declining to the present level around 1.0.

Figure 15 also shows the ratio of NII to NIM; this ratio hovered around 30% until the early 1990s and the jumped to 100%
by the end of the period of study. The peak was reached, not surprisingly, in 2006 (111%), the year before the beginning of the financial crisis.

Figure 16 shows a scatter diagram of two important indicators of overall profitability: the net interest margin to total asset (vertical axis) and the ratio of net income to net interest margin (horizontal axis). Each point is, therefore, a measure of net income (the product of y and x), which allows us to draw the “iso-curves” of NI (at the levels of 2.4%, 3.2% and 4.0% of total assets, respectively). Three different periods clearly stand out: first, the 1970s and the 1980s, where we had high interest margins and a low non-interest income. The 1990s experience a significant shift to the right, but on a lower iso-margin. Finally, in the last decade, banks reached the lowest iso-margin curve, particularly after the crisis, as a consequence of the strong decline in interest margins. This is despite the horizontal axis being close to 200, which means that non-interest income is near a 1:1 relationship with interest income.

A conjecture that has been formulated in the literature [Albertazzi and Gambacorta (2006)] is that such revenue diversification is good, as it allows banks’ profits to stabilize [Saunders and Walter (1994); Lown et al. (2000)]. On the other hand, other studies stress that traditional intermediation activities remain the core business of most profitable banks in the U.S. [DeYoung and Rice (2004)] or argue that a higher proportion of non-interest income increases the volatility of bank profits [DeYoung and Roland (2001), Stiroh (2004)].

Our analysis shows that Italian banks’ non-interest income grew

\[ \Delta BR_t = \alpha + \gamma (BR_{t-1} + \beta MR_{t-1}) + \sum \beta_i \Delta MR_{t-i} + \epsilon_t \]

<table>
<thead>
<tr>
<th>Loan rate</th>
<th>Deposit rates</th>
<th>Bond rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum \beta$</td>
<td>$\theta$</td>
<td>$\sum \beta$</td>
</tr>
<tr>
<td>1970–2012</td>
<td>0.36</td>
<td>1.18</td>
</tr>
<tr>
<td>1970–1984</td>
<td>0.23</td>
<td>1.28</td>
</tr>
<tr>
<td>1985–2012</td>
<td>0.60</td>
<td>1.09</td>
</tr>
<tr>
<td>1985–1998</td>
<td>0.41</td>
<td>1.09</td>
</tr>
<tr>
<td>1999–2012</td>
<td>0.54</td>
<td>1.09</td>
</tr>
</tbody>
</table>

Table 1: Short- and long-term elasticities estimates of loan, deposits and bond rates to short-term market rates changes
Quarterly data from 1970 to 2012; authors’ estimates.
$\sum \beta$ = instantaneous multiplier; $\theta$ = long-term elasticity.
steadily — perhaps even more than expected — during the last three decades, as shown by the shift to the right on the horizontal axis of Figure 16. This was stimulated by the developments in the financial markets (which created great opportunities for widening the range of services offered to households and companies) rather than by a transition to an investment banking model or a universal banking model [CEPS (2011)]. The net result is that the diversification of revenues took place, but the expected stabilization of profits did not materialize.

5. Operating costs, operating profits and overall efficiency
A large component of the NI of a bank are absorbed by administrative costs (staff and other), which are significant, particularly in the retail business which is at the same time both labor (staff) and capital intensive (branch network and technology).

The operating profit of a bank can be measured both before and after loan loss provisions. Of course, only the latter can be considered a proxy for the overall profitability of the business in a given year, but as data for provisions are available only after 1974, we will initially refer to gross operating profits. This margin, being independent of the cyclical component of credit risks, reflects the basic profitability of the business and is, therefore, a better indicator of economic efficiency than other proxies, such as the cost-income ratio. The gross operating profit (GOP) is shown in Figure 17. GOP has experienced two long-term trends: an increase between 1965 and 1981 (from 0.9 to 1.9) and a decline ever since. These trends are interrupted by sharp cyclical movements with peaks in 1981, 1986, 1993 and 2001. Only 2012 shows a modest rebound. Since 2001, the decline has been constant, and by 2012 GOP had reached 0.63, 33% of the peak level of 1981 and 39% of the levels reached only 10 years earlier. In other words, the decline of gross margins was not counterbalanced by a corresponding decrease in operating costs.

To investigate efficiency issues of the banking system further, we analyze operating cost margins using two traditional indicators of relative efficiency: the ratio of operating costs to total assets (a proxy for the size of the banking system) and the ratio of operating costs to net income (a measure of economic efficiency, the so-called cost-income ratio). As demonstrated in Figure 18, operating costs fluctuated around 3.0 of total assets until the early 1990s, then declined sharply to 1.5.

From this point of view, the gains in efficiency of the last two decades seem remarkable. The starting point was 3.3 in 1965 and in 1990, it stood still at 3.1. But in 10 years, it declined to 2.0, and reached 1.2 in 2012, the lowest level of the entire period. One can, therefore, deduce that the consolidations that took place during this period had apparently allowed for...
significant economies of scale and scope. On the other hand, the other measure of efficiency, the cost-income ratio, does not show such a decline. It fluctuated around 70% until the 1990s and around 60% in the following decades, but in 2012, thanks to the first improvement of an entire decade, it is still at the same level as the early 1980s. Of course, both ratios shown in Figure 17 are very rough proxies of efficiency. Total assets are not the best measure of bank production, and therefore, the ratio of total costs to total assets cannot be considered a sound measure of unit costs. On the other hand, the cost-income ratio is influenced by the size of the denominator, which depends on market factors and in particular on the level of interest rates.

The overall picture that emerges from Figure 18 is that Italian banks have reached efficiency with respect to the scale of activity. However, over the long period of declining margins, operating costs proved rather rigid. As a consequence, as of 2012 operating costs still absorb two-thirds of gross operating profits.

One can, therefore, ask whether there are still areas of inefficiencies in the Italian banking system. The answer cannot be obtained from the aggregate data we have used so far. However, in this particular case an international comparison can help, as operating costs depend also on a series of structural factors for which we have comparisons [ECB (2010)].

If we look at Table 2, we can observe that:

- The Italian banking system has doubled the number of branches since 1990, in sharp contrast to other banking systems. Even if this is due to the late liberalization, the difference is striking. It must also be stressed that based on past history, financial crises are always followed by significant restructurings.
- Looking at the right part of the table, Italy shows significant symptoms of overbanking in relation to three main indicators: branches’ total assets (size of the production unit); GDP (size of the underlying economy); and population. Only Spain looks more overbanked than Italy.
- With the expansion of branches, the number of employees has significantly increased, with negative effect on productivity, as measured by total asset per employee. Italy also has a low number of employees per branch (9.7, higher only than Spain, in contrast with an average of 12.9 for euro area and 14.5 for the E.U.). This seems to point to a wider diffusion of mini-branches, aiming to maximize the density of the branch network.

These result might at first seem surprising, given the strong consolidation of the last decades. Where have the alleged synergies gone? The answer is that mergers in Italy are typically undertaken with the aim of boosting revenues rather than cutting operating costs. Perhaps to soften the opposition in local communities, the strategic plans have always stressed the...
low level of overlapping in the branch networks (and hence in the allocated staff), but particularly in the urban areas (chased by all banks, whatever their size, in the years following the liberalization) the banking presence can now be considered redundant, predominantly because technology has decreased the need for proximity, which has historically been the reason for the high density of branch networks.

It must be stressed that the cost-income ratio of the Italian banks does not seem particularly high when compared with other euro area countries; the only exception is Spain where the ratio dropped thanks to the very high level of gross profits (Figure 19).

6. Provisions, profit before taxes, RoA and RoE

Provisions are obviously a major component of the banking business, so the indicators of profitability considered so far can lead only to rough conclusions about the overall profitability of the banking system. Provisions and losses (always as a percentage of total assets) show wide fluctuations, mainly determined by the economic and credit cycles. As Figure 20 demonstrates, the ratio of provisions to total assets peaked in 1981, the mid 90s and in 2012. These peaks reflect the three main shocks faced by the Italian economy: the increase in oil prices and inflation (with heavy structural consequences for major Italian industries); the devaluation of the lira coupled with the emergence of huge losses in the banks of Southern Italy; and the financial crisis.

It is worth stressing that provisions as a percentage of total assets (and therefore of the loan portfolio) peaked only in 2011 and 2012, a few years after the crisis erupted. Although the slowdown of 2010 could have been caused by an underestimation of the overall risks of the new environment (using provisions as a buffer to smooth the decline of gross profits) the trend shows that in the three years after the crisis, the quality of the Italian banks’ loan portfolios were significantly better than in previous crises. However, the picture changes dramatically in 2011: the duration of the crisis (coming after a long period of sluggish economic growth) coupled with the contagion to Italy of the peripheral Europe difficulties, has brought the ratio of provisions to total assets to a record level for the entire period. But due to the decrease in gross profitability, provisions now absorb more than 70% of net income, more than twice the level of every single year of the entire period.

As a result, the combination of the cycle of operating profits with the cycle of provisions, leads to two rising trends of profits before taxes, abruptly interrupted by the recessions of the 1990s and the current one (Figure 21). Not surprisingly, 2011 and 2012 show negative values. As this measure is the best proxy of banks’ return of assets (Haldane (2009)), we can say that in each period profitability seems to revert to the value of the beginning of the cycle and that the two worst years of the financial crisis (2011 and 2012) have brought overall profitability (RoA) to negative level, for the first time in the recent history of the Italian banking system.8

The financial crisis has led to a decline in RoA by 90 basis point in only five years. As far as the basic profitability is concerned, the financial crisis seems to have once more pushed the RoA back to where it was a few decades ago, even if one considers 2011 figure as exceptional (the 2010 figure is 70 basis points lower than 2007). In other words, the impact of the crisis on overall profitability seems to have been particularly severe in Italy. Figure 21 also shows profits after taxes (PAT). Interestingly enough, the distance between the two curves (the “fiscal wedge”) has narrowed in recent years. PAT is also the best indicator of the net profitability of the business. As the graph shows, the financial crisis marks the end of the second cycle of profitability, with levels in 2009-10 being very close to the troughs of the mid 1970s and mid 1990s, and the negative values in 2011 and 2012 never having been seen before.

Before the crisis, the return on equity (RoE) was considered the best gauge for banks’ profitability. The crisis has shown that a measure that does not take into account the level of implied risk, is fundamentally flawed ([ECB (2009), Haldane (2009)]. For our purposes, the indicator keeps its importance, even if many caveats must be taken into account. First of all, the Italian banking system, unlike others, did not have significant amounts of “hidden risks” due to toxic assets. Second, RoE measures the capacity to reward the shareholders (particularly significant in a completely privatized system) and to generate internally capital resources.

8 The data shown in the graph do not include the impairment of intangible assets. Including the significant devaluation of the high goodwill embedded in the mergers during the long concentration phase, in 2011 the aggregate value of RoA was -0.6%.
Figure 22 shows the RoE (expressed as the ratio of PAT to capital and reserves) and RoA (after taxes) since 1974. The former shows significant fluctuations over the period. Also in this case, the troughs are determined by the crises of the early 1990s and the current one. It is worth stressing that in the four years prior to the crisis, the average level of RoE was around 10%, quite low based on international standards (even if one has to take into account the high level of Italian taxation) and not far from the levels of the 1980s. Moreover, the average for the years since the crisis (2008-12) is close to zero, with negative values registered in the last three years. The levels reached in the last three years is only slightly above the levels reached during the crisis of the 1990s.

It is clear that RoE and RoA have become decoupled, mainly because the leverage of the Italian banking system decreased substantially over this period. Unlike other financial systems, and in particular the U.S. and the U.K. [Alessandri and Haldane (2009)], leverage decreased significantly in Italy, from 50 to 15, predominantly due to the regulatory pressures from the Bank of Italy to increase the robustness of the banking system. Not surprisingly, the results for RoE have not been as outstanding as in other countries where banks were allowed to significantly increase their financial risk.

It is interesting to decompose the RoE into a profitability effect (RoA) and a leverage effect, according to the usual proxies. The result are shown in Figure 23, where changes in the RoE (net profit) have been decomposed into four elements: RoA, leverage,
provisions and taxation.\(^9\) The periods reflect the main phases of the RoE cycle. We find that leverage contributes negatively to RoE in all periods, but it appears that the more pronounced movements are determined by provisions and surprisingly (particularly in the fourth period (1998–2007) positively by taxation. The latter seems to be attributed to the tax effect of provisions. It is also interesting to stress that changes in RoA were never the most important factor. However, since changes in RoA normally have the same sign as provisions (with the exception of the first period), the net effect on RoE is normally amplified. Bearing in mind that the markdown of the deposit rates no longer gives a positive contribution to profitability, the net return of the loan portfolio is now the main driver of the return to shareholders.

7. A look at disaggregated data

Our analysis is based on the hypothesis that the aggregate profitability of a banking system is an important indicator of the general trend of the market and, therefore, a main driver (or at least a constraint) for individual banks’ profitability. However, given that in the last two decades the Italian banking system underwent an important process of restructuring and consolidation, it is worth looking also at disaggregated data, drawing on Prometeia’s database on main groups of banks by asset size (Table A1 in appendix). The data are comparable only to a limited extent to those used so far, as ours are unconsolidated and refer to units operating in Italy, while Prometeia’s are consolidated for individual banks (including in particular subsidiaries operating abroad). Overall, Figure 24 shows a similar pattern for all of the main indicators. This confirms our hypothesis about the importance of the average trend and contradicts some previous empirical studies that stressed the diverging performance of large groups [De Vincenzo and Quagliariello (2005)], and therefore, the beneficial impact of the consolidation process.

---

\(^9\) The RoE (net profit/equity) decomposition is based on a formula that multiplies the following four factors: RoA\(^*\) = operating profit/total assets; leverage = total assets/equity; provisions = profit before taxes/operating profit; taxation = net profit/profit before taxes. In order to identify the contribution of each component to RoE, their changes were calculated using the formula 
\[
d(R_t) = (R_t - R_{t-1}) / (R_t - 1) \times 100,
\]
where \(R_t\) is the ratio value in period \(t\).
Figure 24: Main performance indicators - average of individual banks
Source: Prometeia, Analisi dei Bilanci Bancari; see Table A.1 in appendix for more details; banks are distributed on the basis of the classification included in the annual report of Bank of Italy of each year.
The following points are worth stressing:

- The decline in gross margins (both NIM and NI) is more pronounced for large banks than others. This can be attributed to the less competitive pressures felt in local markets.
- The decline of the ratio of operating costs to total assets is more pronounced for large banks, pointing to economies of scale in the consolidation process.
- The net result of the previous two points is that the difference in terms of gross profits is less pronounced than other profitability indicators. It seems that both large and small banks have the lowest levels of gross profits at the end of 2012. As small banks were the most profitable at the beginning of the period (50% more than large banks) this group seems to have been particularly hit by the evolution of the last two decades.
- In terms of operating efficiency, as measured by the cost–income ratio, the large banks have experienced the greatest improvements. The high cost-income ratios of the early 1990s can be attributed to the very low levels of profitability of a few major large banks that were later acquired by more efficient banks. The sharp decline of the cost-income ratio of large banks in the following years is, therefore, proof of the success of consolidation and of the internal solution given to the problem banks of those years. Profits of the past were somehow used to cushion against the crises of the early 1990s that most heavily impacted banks based in southern Italy. In more recent years, the four classes show very similar levels and trends of cost-income ratios, but it must be stressed that at the end the highest ratios can be found for the two extreme classes: large and small banks.
- The same ranking, albeit with smaller differences, can be seen for RoA. In this case, big and medium banks show the highest values. But for all groups, the 2012 value is the lowest of the last 16 years.
- RoE levels had been quite similar (and with similar patterns) until the financial crisis. The dispersion around the average for this series is less pronounced than for any other series. This common pattern was broken by the financial crisis. Only small and medium banks could maintain positive values for RoE after the crisis, large and big banks turned into red in 2010 and 2012.

8. Conclusions

The Italian banking system is typically retail-orientated. In a recent study [CEPS (2011)], it was found that even the two largest banks in the country fell into the retail cluster, while many other countries’ major banks were classified either as universal or investment banks. Moreover, it cannot be said that Italian banks expanded “beyond their economic and social value” [Llewellin (2012b)]. Although since the mid 1990s, the level of bank activity grew to record levels, the size of the banking system, as measured by the ratio of total assets to GDP, is still significantly below the European average.

These were the main reasons why at the beginning of the crisis the Italian banks appeared to be better off than their peers from most other countries. However, when the crisis hit the periphery of Europe, the scenario changed abruptly and the points of relative strength deteriorated rather quickly. Consequently, even for a retail-orientated banking system, the crisis will prove to be “transformational” [Llewellin (2012a)].

At least for the first two decades of the near half century covered by our research, the Italian banking system was strictly regulated and enjoyed the typical contradictions of a protected market: it was barely competitive, essentially domestic and far from efficient [Enria-Focarelli-Landi (1999)]. On the other hand, it was also very profitable, in terms of net income, even though a large part of the gross profits was absorbed by operating costs. The combination of high margins and high costs led to returns on assets that were not significantly different from other banking systems. In a nutshell, the quasi-rents, made possible by oligopolistic conditions created by a strict regulation, were absorbed by extra-costs (more administrative costs, more personnel, higher average pay or a combination of the three things) rather than being passed through to net profits.

The long phase of structural reforms, financial innovation and globalization, ending with the monetary union has profoundly changed the landscape and brought mixed results. Margins declined steadily, as a consequence of the decline in interest rates, reaching their minimum levels by the end of the period of our study. One can argue that, thanks to increased competition, the oligopolistic component of profits declined. However, our data does not allow for separating the different effects of changing market conditions on the overall profitability.
From the point of view of revenues, Italian banks proved to be very quick in moving from interest income to non-interest income. However, the latter declined (albeit at a slower pace) along with the former, so that in the end both net interest margin and net income (as a percentage of total assets) were at their lowest levels over the entire period of study and less than 50% of the peak reached in the late 1980s and early 1990s. In two decades, the gross profitability of the industry halved.

The long-term decline in the profitability of the traditional intermediation business (as reflected by NIM) is the main finding to be stressed and seems to be the Achilles’ heel of the Italian banking system today. In particular, our econometric exercises have shown a significant elasticity of interest margins to market rates, which is very bad news in a long period of declining interest rates. On the one hand, the deposit component of the spread became even negative when rates approached the zero bound while the sensitivity of the loan component was heightened by the large diffusion in Italy of short-term loans or variable rate long-term loans. This preference is due both to demand and supply factors. At any rate, in the Italian banking system the duration of the loan portfolio is very low and banks are exposed (more than in other countries) to the downside effects of declining interest rates.

Moreover, loans – particularly since the advent of the euro – grew much faster than deposits, creating a large funding gap, and therefore exposing net interest margins to the fluctuations of market interest rates, which have been significant since the start of the financial crisis.

The problem was barely noticed during the initial years following the monetary union, because of the “anesthetic effect” coming from two factors: the capital gains on the government bond portfolio and the high level of fees that could be extracted in a period of booming markets (commissions on asset management being the most important example). Particularly in the 1990s, but also for a period afterward, the portfolio of government securities was a sort of “dowry” that benefited the profit and loss account, but disappeared thereafter.

The crisis has completely changed the landscape and, in a sense, has put the engine into reverse gear: the funding costs soared, the government bond portfolio generated losses instead of profits and the level of commissions was kept down by the low level of economic activity and the depressed conditions in the financial markets. Italian banks (unlike their international peers) did not record “fake” profits coming from the bubble conditions of the credit markets (something which is close to selling out-of-the-market options) but probably underestimated the cyclicality of some components of non-interest income.

It must be stressed that our analysis ends with the exceptional conditions of 2011 and 2012. In those years, Italian banks recorded significant losses on the trading portfolio of government bonds (particularly in 2011) and a significant increase in their funding costs, partially compensated by the wide use of ECB lending facilities. The net effect is that both net interest income and net income significantly declined in comparison to the previous years.

The silver lining of the present decline in margins is that the previous years’ depressed level of non-interest income also has a significant cyclical component which will benefit the P&L in the next phase, when, hopefully, the spread on Italian government bonds will be brought down, as happened in 2012 in comparison with the previous year. From this point of view, the destiny of Italian public finances and of Italian banks are strongly connected. In case of a happy ending to the European crisis, the government bonds bought by Italian banks in recent years could prove to be a new “dowry” for future profitability, as it happened in the 1980s and in the 1990s.

However, this does not change the fact that the level of NIM, which is the primary source of Italian banks’ profitability, has converged toward the levels of other countries, even if it is still higher than others (Spain being now the only system with a significantly higher NIM). It seems very unlikely that NIM could increase in the future, even in the favorable scenario of a quick solution of the European crisis. In the foreseeable future, central banks will continue to keep market interest rates at low levels and banks now have no incentives to change their exposure to the interest rate cycle. Moreover, the loan portfolio will not increase faster than the nominal GDP. In fact, it is declining due to the ongoing credit crunch. Apart from the demand factors, there are strong supply constraints coming from the new regulatory framework (in particular Basel 3 capital and liquidity ratios) and the need to reduce the funding gap. In other words, neither prices nor quantities are likely to give a significant boost to future profitability in terms of NIM.
If Italian banks want to increase their profitability, they must therefore concentrate on the items below net income, i.e., operating costs and provisions. The former has proven itself to be rather rigid so far. Operating costs decreased as a percentage of total assets, but mainly as a consequence of an increase in the level of activity (i.e., the denominator). Not surprisingly, the rigidity has been even stronger in the years following the crisis. This seems to be one of the main points to be underlined, particularly because international comparisons show that on many accounts Italian banks are still burdened by labor and structural costs linked to branch networks that are much larger than in other countries. Our data show that the benefits of the massive consolidations of the past decades are far from being fully delivered.

The second component that will matter in the future is provisions for expected loan losses, which are outside the control of banks and basically follow the economic cycle. Our analysis has shown that during the last decade the ratio of provisions to total revenues fell to historically low levels and then soared in 2011 and 2012. The low levels of provisions in the first couple of years after the crisis (particularly in 2010) can be attributed partly to a high quality of the loan portfolio in comparison to previous crises and by special measures in support of the corporate sector (the moratorium on a large part of the loans agreed upon by the Italian Bankers Association). It seems quite likely that at the beginning of the crisis Italian banks tried to use provisions to cushion against falling profits, hoping for an immediate rebound of the world economy. Notwithstanding the sharp correction of 2011 and 2012, there is still an implicit risk in Italian banks’ loan portfolios, which could have a negative impact on future profitability. Consequently, the pressure to cut operating costs will be even higher.

Our analysis shows that in the past 45 years, the RoA (before taxes) of Italian banks has followed two wide cycles, with the 2010 level very close to the troughs of the mid 1970s and late-1990s. The 2011 and 2012 figures have been even negative. The crisis has also led to a significant devaluation of the high goodwill embedded in the mergers during the long concentration phase. The effect has been striking in 2011 and 2012 (and in particular explains the heavy losses of 2011), but in the present uncertain scenario could also continue. The beginning of the second decade of the new millennium can, therefore, be considered one of the bleakest periods in the history of Italian banks. For the reasons explained above, a new positive cycle can be driven only by a significant decrease in operating costs.

This is particularly important from the viewpoint of the implications for future profitability in terms of RoE. Our analysis shows that Italian banks’ RoE has been traditionally lower than their peers in other countries. As the level of capital has increased over the years, leverage gave a negative contribution to the RoE. The only positive and significant contribution came from taxation. This seems to be paradoxical as Italian banks have always complained about a tax burden significantly higher than their competitors. As a matter of fact, it is because of these high tax rates (and the unfavorable regime of taxation for loan losses) that changes to them can have significant effect on net profits. For example, in the late 1990s, the new tax regime for loan loss provisions boosted profitability. Until the financial crisis, the net effect was an increase in RoE, to average levels that were still lower than in other countries but sufficient to attract capital, to boost equity prices and make the mergers of those years look very profitable.

Overall, for Italian banks, RoA has been the main driver of RoE (taking into account provisions for loan losses). Consequently, RoE also experienced two cycles (albeit less pronounced than RoA). Even without taking into account the negative values of 2011 and 2012, the level of 2010 (2.99) is close to historical lows and very far from the levels reached in 2006 (11.44). It is also interesting to note that the highest level was reached in 1986 (11.73), which seems to have been the year when the “golden age” of high margins came to an end.

In the U.K., the banks’ high RoE have been proven to be a “mirage” more than a “miracle” [Haldane (2009)], the distortion being caused by the increasing leverage, higher than expected level of risk implied in the overall portfolio and by implicit subsidies [Llewellyn (2012b)]. In Italy, as we have seen, the distortion was mainly due to the cyclical components of RoA. This created a sort of myopia that led banks to take as permanent changes in profitability that were either conditional on the interest rate cycle or dependent on one-off changes in the taxation regime. With the notable exception of the trend in loss provisions, these factors are not comparable with the misrepresentation of risks that many believe is one of the main reasons for the “mirage” of bank profitability in the last decades.
Our analysis also explains why the financial crisis has so quickly changed the prospects of Italian banks. At the beginning, they were more robust than most banking systems, both from the standpoint of profitability (particularly in terms of RoA) and capitalization. The former declined very quickly, mainly because of the high sensitivity of margins (particularly through funding costs) to market rates, the deceleration in the rate of growth of intermediation and the increasing burden of loan loss provisions. On the other hand, the worsening of the crisis led the market (and the regulators) to ask for higher levels of capital.

The Italian banks’ business model, strongly concentrated on traditional intermediation, and their capital levels was a blessing when the financial crisis hit, and made Italian banks very different from others that were heavily exposed to assets whose value deteriorated rapidly (the so-called “toxic” assets). It is important to stress the strengths of the Italian banking model: traditional intermediation and prudent management (the latter to be ascribed also to the effectiveness of prudential supervision), since in other countries even banks that were dedicated to lending (the British building societies, the Spanish Cajas, the German Landesbanken) had significant difficulties when the crisis hit.

The situation of Italian banks deteriorated, and more quickly than expected, as soon as the crisis hit the government debt market, causing an abrupt rise in funding costs and so reducing an already depressed net interest margin. By the same token, the level of capital which was high, but not exceptional, at the beginning of the crisis proved to be inadequate when the deterioration of the situation led markets, the European Council and then the European Banking Authority to recommend a further injection of capital. Notwithstanding very difficult market conditions, the first (and biggest) rights issue by an Italian bank (Unicredit) has been successful (Saccomanni 2012), and overall the entire recapitalization after the November stress-tests proved to be necessary to avoid further problems to euro area banks [EBA (2012)].

This success must also be credited to the European Central Bank, which has injected liquidity in exceptional amounts and at exceptionally low interest rates. At the beginning of 2012, Italian banks were reported to have tapped the end-2011 LTRO for €116b, with ECB refinancing representing at the time 25% of total bank bonds outstanding, or 10% of total deposits [Barclays Capital (2012)]. In other words, the present profitability of Italian banks has been sustained to a large extent (and artificially) by the extreme generosity of the European Central Bank. This will add to the pressure for a “natural” rebound of profitability, as soon as normal market conditions are restored.

All this leaves us with a major question. Is the Italian banking system a fallen angel? [Resti (1997)] Our analysis has shown the profound changes of the last 50 years and allows us to conclude that in a sense the long-term factors behind the high profits of the past have been exhausted. In the first phase, the high profitability was explained by the low competition of Italian retail bank markets (protected by heavy regulation); afterward, it was the long phase of declining interest rates and the strong increase in the overall levels of activity that had a positive effect. Both of these factors are now over.

Despite the fact that the financial crisis did not have catastrophic effects, it has brought banks’ RoA near the troughs of the previous profit cycles, even without considering the negative values of the last two years. While other banking systems are simply “climbing out of the holes they had dug for themselves” – as it has been said for the U.S. banking system [Mayo (2011)] – the problem for Italian banks is the continuous erosion of the basic source of profitability. The crisis has compounded its effects on problems that were already visible, rather than introducing an abrupt change in terms of unforeseen and unaccounted losses. So the question can be rephrased: what will be, if any, the main driver of a future positive trend in the profitability of Italian banks? Our analysis suggests that neither prices (rates) nor quantities (the level of activity) are likely to make a positive contribution. Consequently, a significant reduction in operating costs appears to be the only way to bring RoA and RoE to “normal” levels. The good news is that the present high level of operating costs gives ample room for future cuts. The bad news is that the “idyllic” environment of the past is over. Particularly in the decades before the financial crisis, banking in Italy looked like a “Pareto paradise” where all banks’ stakeholders were more or less happy: clients, employees and shareholders. In the future, Italian banks will have to choose and make tough and unpalatable decisions. Angels perhaps are still flying, but it is the Paradise that no longer exists.
A long-term approach to Italian banks’ profitability: paradise lost?

Appendix

<table>
<thead>
<tr>
<th>Number of banks</th>
<th>59</th>
<th>111</th>
<th>116</th>
<th>65 (IAS compliant)</th>
<th>116 (IAS compliant)</th>
<th>112 (IAS compliant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>4.8</td>
<td>3.9</td>
<td>2.7</td>
<td>2.8</td>
<td>2.3</td>
<td>2.0</td>
</tr>
<tr>
<td>1995</td>
<td>6.7</td>
<td>5.3</td>
<td>5.4</td>
<td>4.9</td>
<td>4.2</td>
<td>4.2</td>
</tr>
<tr>
<td>2000</td>
<td>4.0</td>
<td>4.0</td>
<td>3.7</td>
<td>3.0</td>
<td>3.1</td>
<td>3.7</td>
</tr>
<tr>
<td>2005</td>
<td>2.7</td>
<td>1.4</td>
<td>1.9</td>
<td>1.3</td>
<td>0.5</td>
<td>~0.3</td>
</tr>
<tr>
<td>2010</td>
<td>59.9</td>
<td>76.1</td>
<td>68.4</td>
<td>62.1</td>
<td>74.2</td>
<td>87.4</td>
</tr>
<tr>
<td>2012</td>
<td>NA</td>
<td>5.3</td>
<td>9.9</td>
<td>9.1</td>
<td>2.6</td>
<td>~0.6</td>
</tr>
<tr>
<td>1990</td>
<td>1.3</td>
<td>1.4</td>
<td>1.7</td>
<td>1.4</td>
<td>0.5</td>
<td>~0.5</td>
</tr>
<tr>
<td>1995</td>
<td>20.9</td>
<td>18.4</td>
<td>13.7</td>
<td>12.4</td>
<td>11.6</td>
<td>13.2</td>
</tr>
<tr>
<td>2000</td>
<td>4.3</td>
<td>3.2</td>
<td>2.5</td>
<td>2.9</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>2005</td>
<td>6.1</td>
<td>4.5</td>
<td>5.2</td>
<td>5.1</td>
<td>3.5</td>
<td>3.5</td>
</tr>
<tr>
<td>2010</td>
<td>3.9</td>
<td>3.6</td>
<td>3.2</td>
<td>3.3</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>2012</td>
<td>2.2</td>
<td>0.9</td>
<td>2.1</td>
<td>1.6</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>1990</td>
<td>63.6</td>
<td>80.6</td>
<td>61.9</td>
<td>63.7</td>
<td>67.4</td>
<td>64.7</td>
</tr>
<tr>
<td>1995</td>
<td>NA</td>
<td>3.8</td>
<td>13.0</td>
<td>10.6</td>
<td>3.6</td>
<td>~3.0</td>
</tr>
<tr>
<td>2000</td>
<td>NA</td>
<td>0.9</td>
<td>2.0</td>
<td>1.5</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>2005</td>
<td>17.4</td>
<td>17.9</td>
<td>11.1</td>
<td>11.7</td>
<td>10.9</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Appendix: Table

<table>
<thead>
<tr>
<th>Number of banks</th>
<th>Very large</th>
<th>Large</th>
</tr>
</thead>
<tbody>
<tr>
<td>1990</td>
<td>3.8</td>
<td>4.3</td>
</tr>
<tr>
<td>1995</td>
<td>5.7</td>
<td>6.1</td>
</tr>
<tr>
<td>2000</td>
<td>3.6</td>
<td>3.9</td>
</tr>
<tr>
<td>2005</td>
<td>2.0</td>
<td>2.2</td>
</tr>
<tr>
<td>2010</td>
<td>64.1</td>
<td>63.6</td>
</tr>
<tr>
<td>2012</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

Appendix: Table
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medium</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net interest margin/Total assets</td>
<td>5.1</td>
<td>3.8</td>
<td>2.5</td>
<td>2.4</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Net income/Total assets</td>
<td>7.1</td>
<td>5.2</td>
<td>5.4</td>
<td>4.3</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>Costs/Total assets</td>
<td>4.4</td>
<td>3.8</td>
<td>3.6</td>
<td>2.7</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Gross operating profit/Total assets</td>
<td>2.7</td>
<td>1.5</td>
<td>2.1</td>
<td>0.9</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Cost/Income</td>
<td>62.1</td>
<td>73.7</td>
<td>66.1</td>
<td>62.1</td>
<td>67.9</td>
<td>68.8</td>
</tr>
<tr>
<td>RoE</td>
<td>NA</td>
<td>7.1</td>
<td>11.4</td>
<td>8.8</td>
<td>4.6</td>
<td>-1.1</td>
</tr>
<tr>
<td>RoA</td>
<td>NA</td>
<td>1.4</td>
<td>1.8</td>
<td>1.2</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>Leverage</td>
<td>14.6</td>
<td>13.1</td>
<td>10.7</td>
<td>11.2</td>
<td>10.8</td>
<td>10.6</td>
</tr>
<tr>
<td><strong>Small</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net interest margin/Total assets</td>
<td>5.2</td>
<td>4.3</td>
<td>3.0</td>
<td>3.0</td>
<td>2.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Net income/Total assets</td>
<td>7.1</td>
<td>5.8</td>
<td>5.6</td>
<td>5.1</td>
<td>4.3</td>
<td>4.0</td>
</tr>
<tr>
<td>Costs/Total assets</td>
<td>4.1</td>
<td>4.3</td>
<td>4.0</td>
<td>3.2</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>Gross operating profit/Total assets</td>
<td>3.1</td>
<td>1.6</td>
<td>1.8</td>
<td>1.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Cost/Income</td>
<td>57.0</td>
<td>74.6</td>
<td>70.2</td>
<td>61.8</td>
<td>75.9</td>
<td>68.6</td>
</tr>
<tr>
<td>RoE</td>
<td>NA</td>
<td>5.1</td>
<td>8.5</td>
<td>8.6</td>
<td>2.1</td>
<td>-0.6</td>
</tr>
<tr>
<td>RoA</td>
<td>NA</td>
<td>1.5</td>
<td>1.7</td>
<td>1.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Leverage</td>
<td>12.7</td>
<td>11.4</td>
<td>9.4</td>
<td>10.1</td>
<td>10.9</td>
<td>12.2</td>
</tr>
</tbody>
</table>

Table A1: Main performance indicators - average of individual banks
Source: Prometeia, Analisi dei Bilanci Bancari; banks are distributed on the basis of the classification included in the annual report of Bank of Italy of each year.

Notes: RoE = Profit after taxes/Capital and reserves; RoA = Gross operating profits/Total assets; Leverage = Total assets/Capital and reserves; some not significant values could be excluded from the average.
The introduction of market-based pricing in corporate lending

Ivan Ivanov
Financial Economist, Office of Financial Intermediaries, Division of Economic and Risk Analysis, United States Securities and Exchange Commission

João A. C. Santos
Vice President and Function Head, Financial Intermediation Function, Federal Reserve Bank of New York and Full Professor, Nova School of Business and Economics

Thu Vo
Ph.D. Candidate in Finance, Simon School of Business, University of Rochester

Abstract
We describe a recent innovation in the corporate lending business whereby banks tie the interest rate during the life of the loan to the borrowers’ credit default swap spreads or to a CDS index. We also discuss the potential impact this innovation may have on bank lending and more generally on financial intermediation and identify some potential adverse effects it may have for the stability of the financial system.

1 We thank James Traina and Phoebe White for outstanding research assistance. The views stated herein are those of the authors and are not necessarily the views of the Federal Reserve Bank of New York, the Federal Reserve System, the Securities and Exchange Commission or of Ivanov’s colleagues upon the staff of the Securities and Exchange Commission. The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees.
1. Introduction

Bank lending has been synonymous with monitoring, private information and relationship borrowing. For instance, the typical bank loan up to the late 1980s used to carry a fixed interest spread over a base rate, such as prime or LIBOR. Even though borrower credit quality is likely to change after loan origination, fixed interest rate spreads are not much of a problem for banks because of the presence of financial covenants in the loan contract. Financial covenants are set such that a sufficiently large credit quality deterioration triggers a covenant violation, giving the lender the right to accelerate the maturity of the loan.

Lenders almost never call loan repayment because of existing relationships with borrowers, instead they most often force renegotiation and modify interest rate spreads to reflect the underlying credit quality of the firm [see Smith and Warner (1979)]. On the other hand, if credit quality improves, borrowers initiate renegotiation with their lender to reduce interest rate spreads, incurring economically significant renegotiation fees ranging from 10 to 40 basis points of loan amount [see Dennis and Mullineaux (2000)].

Financial deregulation in the 1980s and 1990s intensified banking competition and spurred financial innovation that challenged the traditional bank lending model. Banks started including performance-based pricing features in loans. Performance pricing grids tie loan spreads to measures of the firm's financial health, such as credit ratings or leverage ratios. For example, Table 1 presents the pricing grid from one of the loans in the sample employed in Ivanov (2012). This pricing grid is tied to the adjusted leverage ratio of the borrower and has four pricing steps. At the low rate end of the grid, the borrower pays an interest spread of 100 basis points over LIBOR whenever the adjusted leverage is between 3.5 and 4. If the financial condition of the firm deteriorates and the leverage ratio exceeds 5, the firm pays 250 basis points over LIBOR.

Ivanov (2012) reports that the use of performance-based pricing contracts became widespread in the early 1990s and that approximately half of bank loans nowadays include these pricing provisions. He finds that the primary role of performance pricing grids is to delay costly renegotiation, reducing the probability of interest spread-decreasing renegotiation outcomes, while having little effect in case of borrower financial health deterioration. The author also argues that banks employ pricing grids as a marketing tool to stay competitive with borrowers.

As competition continued to grow, banks looked for alternative ways to innovate their lending business. This led banks to the Credit Default Swap (CDS) market, which provided them with a unique opportunity to employ borrower-specific information contained in credit default swap rates in bank lending activity. In fact, since 2008 banks have increasingly extended loans to corporations with interest rate spreads tied to the borrower's credit default swap spreads. This financial innovation is often referred to as market-based pricing and raises fundamental questions about the future of the current monitoring-based bank lending model. In the next section we describe the institutional background related to market based-pricing. The last section concludes, discussing some implications of this innovation for bank monitoring.

2. Institutional background

Market-based pricing ties loan interest rate spreads to borrowers' CDS spreads or to a Credit Default Swap index (CDX). CDS/CDX-priced loans first appeared in the second quarter of 2008. Even though performance-based pricing adjusts interest rates spreads to reflect changes in credit quality, a drawback with this feature is

<table>
<thead>
<tr>
<th>Category</th>
<th>Adjusted leverage ratio</th>
<th>LIBOR spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&gt; 5.00 to 1.00</td>
<td>2.50%</td>
</tr>
<tr>
<td>2</td>
<td>&gt; 4.50 to 1.00 but ≤ 5.00 to 1.00</td>
<td>2.00%</td>
</tr>
<tr>
<td>3</td>
<td>&gt;4.00 to 1.00 but ≤ 4.50 to 1.00</td>
<td>1.50%</td>
</tr>
<tr>
<td>4</td>
<td>&gt;3.50 to 1.00 but ≤ 4.00 to 1.00</td>
<td>1.00%</td>
</tr>
</tbody>
</table>

U.S.$325 million credit agreement of Pharmerica Inc. (31 December 1998)

Table 1: Performance-based pricing

This table presents an example of a performance pricing contract from the private credit agreement of Pharmerica Inc. (31 December 1998). The LIBOR spread is tied to the firm's adjusted leverage ratio. The pricing grid maps the adjusted leverage ratio onto an interest rate spread for each pricing grid step.

---

that time-varying risk premia make the fixed spreads at each step of the pricing grid inaccurate. Market-based pricing circumvents this problem because CDS spreads are affected both by borrower financial health and time-varying risk premia.

Market-based pricing was advertised in the popular press as a valuable arrangement to ease banks’ resistance to lend at fixed interest rates, given the highly uncertain economic conditions at the time. A 1st July 2008 article in the Dow Jones Newswire stated, while referring to banks’ granting of credit lines: “The borrower is ensured continued access to capital through market cycles, and the lender, if and when the facility is drawn upon, gets a market-based price…”

Market-based pricing contracts, in addition to tying the loan interest rate spread to the borrower’s CDS spread, often specify either an interest rate cap or a floor, or both, that are usually kept constant throughout the life of the loan. If the CDS/CDX stays within the floor-cap range, the borrower pays an interest rate spread on its loan that varies with the CDS/CDX. If the CDS/CDX spread exceeds the cap, the borrower pays the interest rate cap, and whenever the CDS/CDX spread is below the interest rate floor, the borrower pays the interest rate floor. Intuitively, the interest rate cap benefits borrowers, while the interest rate floor benefits lenders.

Some loan contracts tied to CDS/CDX allow the interest rate floor and/or cap to vary throughout the life of the loan. These contracts almost always tie the interest rate floor and cap to the credit rating of the borrower. Table 2 presents an example of one of these contracts from the sample of loans employed by Ivanov et al. (2013). According to the contract, if the borrower retains a rating above AA-, the floor is 20 basis points, while the cap is 87.5 basis points. As long as the borrower’s CDS spread stays within the 20bps–87.5bps interval, the borrower pays an interest rate spread linked to the CDS spread according to a formula specified in the contract. If the borrower’s financial condition deteriorates and it is downgraded to say A or A+, then the floor will go up to 25 bps and the cap will increase to 100 bps. Again, after the downgrade as long as the borrower’s CDS spread stays within the 25bps–100bps interval, the borrower pays an interest rate that evolves with the CDS spread according to the formula specified in the loan agreement.

Ivanov et al. (2013) present evidence that the interest rate cap is not binding most of the time. In contrast, borrowers hit the interest rate floor relatively often. The authors argue that this is suggestive either of lender bargaining power, and as a result non-competitive pricing, or of insufficient information in CDS spreads.

Figure 1 shows that thus far banks have used the market-based

<table>
<thead>
<tr>
<th>Category</th>
<th>Senior rating</th>
<th>Floor</th>
<th>Cap</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AA≤X≤AAA</td>
<td>20 bps</td>
<td>87.5 bps</td>
</tr>
<tr>
<td>2</td>
<td>A≤X&lt;AA–</td>
<td>25 bps</td>
<td>100 bps</td>
</tr>
<tr>
<td>3</td>
<td>A–≤X&lt;A</td>
<td>35 bps</td>
<td>112.5 bps</td>
</tr>
<tr>
<td>4</td>
<td>X&lt;A–</td>
<td>50 bps</td>
<td>125 bps</td>
</tr>
</tbody>
</table>

U.S.$1.5b five-year revolver of 3M Co (August 05, 2011) LIBOR margin is tied to company’s senior unsecured LTD ratings by S&P and Moody’s and one-year CDS midrate

Table 2: Market-based pricing

This figure illustrates an example of a market-based pricing contract for 3M Co. on 5 August 2011. The LIBOR spread is tied to firm’s one-year CDS spread. The contract has a pricing grid, which specifies floor spread and cap spread at each interval of firm’s credit ratings. Whenever the spread exceeds the cap, the firm pays the interest rate cap; whenever the spread is below the interest rate floor, the borrower pays the interest rate floor.

Figure 1: MBP loans by credit rating

This figure plots the number and total amount, in billion dollars, of market-based pricing loans by S&P credit ratings for the period from the second quarter of 2008 to 2012. This includes all facilities tied to CDS/CDX from DEALSCAN, before matching the DEALSCAN sample with COMPUSTAT and other bank-related databases.
Banks have used market-based pricing to extend loans predominantly to large corporations. Even though the number of CDS/CDX-price based loans issued per quarter rarely exceeds 20, the amount of total debt issued under these contracts has been as high as U.S.$95b per quarter. The focus on large corporations is to be expected since the CDS market is dominated by large corporate borrowers.

### 3. Implications

When market-based pricing first appeared, the popular press focused on the conjecture that this way of setting loan interest rates is valuable because it could protect lenders against unexpected future changes in borrowers’ financial condition. However, an alternative explanation for this financial innovation is that banks started using market-based pricing due to competitive pressures. Since monitoring becomes less important when banks tie loan spreads to borrowers’ CDS spreads, banks can save on monitoring costs when they rely on market-based pricing.

Ivanov et al. (2013) investigate the effect of market-based pricing on the cost of bank credit for corporations. They find that borrowers save on interest spreads both at the time of loan origination, and during the life of the loan by choosing market-based pricing contracts. Their study shows that these findings are robust to controlling for credit quality and selection problems. These findings are consistent with either one of the two hypotheses put forth above. However, using results from additional tests, the authors argue that it is the reduction in monitoring costs, rather than the additional protection these contracts offer banks, that drives the decline in the cost of bank credit. For example, they document that market-based pricing contracts are associated with a much simpler covenant structure, suggesting that the interest cost savings of borrowers are explained by reduction in lender monitoring costs and in expected renegotiation costs due to covenant violation. Thus, market-based pricing appears to differ from performance pricing in that it addresses credit quality deteriorations. For example, Ivanov (2012) argues that his empirical tests suggest that the primary purpose of performance pricing is to decrease bank loan renegotiation only when borrower credit quality improves.

Ivanov et al. (2013) argue that association between market-based pricing and covenant structure could have several important effects on financial markets. For instance, market-based pricing could reduce the benefits of relationship lending, such as intertemporal interest rate smoothing or the complementarity between banks and non-bank financial intermediaries. For example, Boot (2000) argues that banking relationships could accommodate intertemporal smoothing because banks could be compensated for losses incurred early in the relationship with gains that come later. Market-based pricing has the potential to decrease the importance of such inter-period subsidies, even though interest rate caps and floors act as counterbalancing forces.

Another benefit of monitoring is that it could facilitate non-bank financial intermediation. Diamond (1991) and Hoshi et al. (1993) show that other capital market participants use the borrower-specific private information generated by monitoring and that this eases financing from non-bank sources. Since the information contained in CDS spreads is only an imperfect substitute for the information produced by bank monitoring, the widespread use of market-based pricing could lead to an increase in the cost of non-bank funding sources.
Last, Ivanov et al. (2013) explain that market-based pricing has the potential to create liquidity spirals in the cost of bank credit. For instance, shocks to the CDS market unrelated to borrower credit quality could lead to increases in CDS spreads. In turn, these increases would make credit for market-based pricing borrowers more expensive. The higher cost of credit would now lead to an increase in borrower CDS spreads and another round of interest rate spread hikes.

4. Conclusion

Even though market-based pricing was first introduced in the second half of 2008, it has become widespread among large corporate borrowers. According to Ivanov et al. (2013), market-based pricing has led to a reduction in the cost of bank credit to corporations. This cost decline does not appear to derive from the additional protection that market-based pricing offers banks against future changes in borrowers’ financial condition. Instead, it appears to derive from the savings in the monitoring costs that banks enjoy when they use this innovation to set interest rates on their loans to corporations. These savings may come with some potential costs. For example, market-based pricing may introduce spirals between CDS spreads and loan interest rates.

References


Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

Matthias Köhler
Department of Financial Stability, Deutsche Bundesbank

Abstract
In this paper, we give a brief overview of how business models developed in banking and outline possible strategies with which banks can restructure their business models in the post-crisis period. We show that large banks with a focus on investment banking and trading activities were not the only entities affected by the crisis and in need of change. Small, retail-oriented banks have to rethink their business models as well, even though they were much less affected by the 2007/2008 financial crisis. The important difference between the two types of banks, however, lies in how their business models need to be changed.

1 The paper represents the author’s personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.
1. Introduction
Banks’ business models have changed considerably over the past few decades. Banks have gone from engaging mostly in traditional commercial banking activities to, over time, becoming increasingly active in non-traditional activities, such as investment banking and proprietary trading. While this has created new opportunities for banks to generate profits, it has also made banks more complex and exposed them to greater income fluctuations. This fact became evident during the 2007/2008 financial crisis, which primarily affected large, investment-oriented banks. Due to these banks’ size and systemic importance, the recently introduced Basel III banking regulations are intended mainly at reducing their riskiness. These regulations include higher capital and liquidity requirements, both of which will reduce banks’ profitability, since their ability to make profits depends much on the extent to which they can leverage their balance sheet and engage in short-term funding. Profitability will likely decrease further if the recommendations of the Liikanen Group (2012) are implemented. This group proposes that banks should move their trading activities to a separate legal entity. These entities could remain within the banking group, but would have to refinance themselves independently. The idea is to prevent banks from using insured deposits to subsidize their trading activities; this should lead to more risk-adequate pricing and reduce margins in the trading business. Overall, therefore, bank profitability appears unlikely to return to pre-crisis levels. For this reason, many banks are under pressure to rethink their business models.

In this paper, we give a brief overview of how business models developed in banking and outline possible strategies with which banks can restructure them in the post-crisis era. We particularly focus on banks located in the European Union (E.U.), although many points made in this paper also apply to banks from other countries. We show that large banks with a focus on investment banking and trading activities are not the only entities affected by the crisis and in need of changing their business models. Small, retail-oriented banks have to rethink their business models as well, even though they were much less affected by the crisis. The important difference between these two types of bank, however, lies in how they have to change their business models.

Analyzing business models is not only important for investors and analysts, but has recently become important for supervisors as well. While the latter were in the past often concerned with capital, liquidity and risk management, the financial crisis has demonstrated that it is also necessary to take a more detailed look at banks’ business models. In general, business models describe how banks generate profits, which customers they serve and which distribution channels they use. Analyzing business models, therefore, goes beyond looking at traditional indicators of bank risk and profitability and should give supervisors a deeper understanding of the sustainability of bank profits and the risks incurred. This allows them to be more forward-looking.

This paper is structured as follows. In the next section, we will briefly describe the main drivers of banks’ business models over the past few decades. In Sections 3 and 4, we broadly distinguish banks according to their business model and how they have performed during the crisis. Possible strategies with which banks can change their business models, and how this may affect the
structure of the E.U. banking sector in the post-crisis era, are outlined in Section 5. Section 6 concludes.

2. Regulatory, technological and strategic drivers of business models

Over the past few decades, the banking sector has become increasingly competitive. This is reflected in the significant fall in the net interest margins over that period. In Germany, for example, the net interest margin has declined from more than 2% in the mid 1990s to 1% in recent years (see Figure 1). To offset this decline, many banks have expanded into other, non-traditional areas of business, which range from the sale of insurance and mutual funds to asset management, underwriting, market making, mergers and acquisitions. These changes indicate that banks have moved away from the traditional commercial banking business model in which a bank makes loans and takes deposits to a more “modern” investment-oriented business model.

Technological advances have been an important driver of increased competition. They have given financial markets and non-banks an advantage over banks and significantly accelerated their growth over the past decades.3 Technological innovations have, for example, made it easier for investment funds to manage their securities and customer accounts. This has lowered the costs of managing a fund and allowed non-banks to offer a broader range of products [Berger (2003)]. Public equity and debt markets have also benefited from innovations in IT, for example, in the area of data handling, experiencing significant reductions in trading costs [Berger (2003)]. Technological innovations have also facilitated the development of securitization markets. Advanced technological tools have made it possible to obtain more accurate pricing and significantly enlarged the set of securitized instruments [Berger (2003)]. All of this has contributed to larger and deeper financial markets and significantly increased the number of products and providers, which might explain why customers have increasingly turned to alternative savings and investment products offered by non-banking institutions.

This is reflected on both sides of banks’ balance sheet. On the liability side, banks have faced an outflow of customer deposits into higher-yielding investments offered by non-bank institutions, such as insurance companies or investment funds. Because it has become more difficult for banks to attract customer deposits, banks have been increasingly turning to wholesale markets for funding. This has not only increased the extent of maturity transformation, but has also allowed banks to increase their lending beyond the growth of their customers deposits and to manage their balance sheet in a more procyclical manner [Adrian and Shin (2010)]. On the asset side, the declining importance of banks’ lending activities is reflected in the ratio of loans to non-banks to total assets, which has decreased significantly over the past few decades. In Germany, for instance, the ratio decreased from about 60% in the 1980s to less than 40% in 2012, while the ratio of deposits from non-banks to total assets declined from 50% to nearly 40%.

Technological innovations have also intensified competition among banks. Innovations such as online banking, for example, have raised transparency and have made it easier for customers to compare different banks’ products. It is also easier for new institutions to enter the market, since new entrants do not need to have a branch network to distribute their products if they use the internet as their distribution channel. This has significantly reduced costs and allowed some banks to offer more attractive interest rates to their customers and to gain market share at the expense of banks with a larger branch network and higher operating costs. Market entry has also been facilitated by the technology of information exchanges, such as credit bureaus [Berger (2003)]. These exchanges collect data from financial institutions, trade creditors, public records and other sources; aggregate and summarize them; and then provide credit reports or credit scores to lending institutions, thus allowing banks to lend to customers with whom they do not have a close relationship. This has substantially lowered the cost of making a loan, since it is no longer necessary for banks to build a close relationship with their customers in order to collect information on their creditworthiness. However, loans that are granted based on hard information, such as credit scores also need to be more standardized, which reduces banks’ scope for differentiating their products and charging higher prices. This suggests that technological innovations have not only made lending more

---

3 Berger (2003) gives different examples of technological changes and examines their economic impact on the banking sector. He finds improvements in costs and lending capacity due to improvements in “back-office” technologies, as well as consumer benefits from improved “front-office” technologies. His research also indicates significant overall productivity increases through improvements in the quality and variety of banking services and shows that technological progress has probably spurred industry consolidation.
Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

competitive, but have also changed the nature of lending, with transactional lending becoming more important relative to relationship lending. This shift may be best illustrated by the evolution of consumer credit, as, for example, is described in White (2007). Combined with greater price sensitivity and reduced customer loyalty, this has reduced the comparative advantage of those banks that focus on relationship banking, because they face higher costs due to their branch network and the large number of employees necessary for their business model.

Financial integration is another important driver of competition in the banking sector. Deregulation and liberalization have caused the banking and financial markets to become more integrated over the past few decades. In Europe, an important step toward an integrated market for financial products was the implementation of the Second Banking Directive in 1989, which introduced a “single passport” for E.U. banks. This passport grants banks located within the E.U. the right to offer services, either through the cross-border provision of services or through the establishment of branches, in any other member state without seeking authorization from the host countries. Another important step toward an integrated market was the introduction of the euro in 1999, since it eliminated currency risk across the euro area and made it easier to take advantage of borrowing and investment opportunities abroad. Because of these and other initiatives, the number and the market share of foreign banks has significantly increased over the past few decades. However, financial integration has not only raised the level of competition at home, but has also made it easier for banks to expand abroad and to offset declining profits in their home markets through higher profits in foreign markets.

The growth of financial markets has also increased competition in traditional lending and deposit markets, forcing banks to look for non-traditional sources of income to make up for the shortfall. The growth of insurance products and mutual funds, for instance, has allowed banks to earn commission income on sales of these products. A number of banks have also set up their own asset management companies in order to generate additional revenues from fund management. Some banks have also started underwriting debt and equity issues, as well as issuing derivatives, and trading securities on their own account on the secondary markets. As a result, banks nowadays offer a much larger range of products and services and are conducting a growing proportion of their activities off-balance sheet (Goddard et al. 2007).

Consequently, while reducing their net interest income, the growth of financial markets has also allowed banks to increase their non-interest income. From a financial stability perspective these changes are positive, since they leave banks with a more diversified income structure and make them less dependent on their home market. However, they have also raised the number of business models and have made banking significantly more complex than in the past.

3. Systematizing business models in banking

Decades ago, banks were mainly focused on domestic markets and commercial banking activities. Nowadays, banks are much more active abroad, and non-traditional activities have become much more important. This has spawned an array of business models. In this section, we give a brief overview of the business models that currently exist in the banking sector. In Table 1, we briefly describe banks’ business models along two main dimensions.

The first dimension distinguishes banks by their products and the customers served. In general, bank customers can be classified as retail (e.g., households and small and medium-sized enterprises) and wholesale customers (e.g., large firms and institutional investors, such as insurance companies and pension funds). As retail and wholesale customers have different financing needs, the products offered differ significantly. For example, while retail customers primarily demand transaction services, savings accounts, mortgages and personal loans, services for wholesale customers are more complex and include wholesale lending, underwriting, market making, consultancy, mergers and acquisitions and fund management. The second dimension, according to which we distinguish banks’ business models is the range of products offered. We call banks that confi

---

4 There are, however, still large differences in the degree of integration across markets. While wholesale markets are already more highly integrated based on price and quantity indicators, retail markets, such as loans to households and small and medium-sized enterprises, lag behind (see various issues of the ECB Report on Financial Integration in Europe). Due to switching costs, differences in fiscal and legal systems and greater information asymmetries, retail-banking markets are still mainly local.

5 For more information on the impact of foreign banks on the retail-banking market in Germany, see Köhler (2008).
themselves to a small range of activities “specialized banks”; such specialization gives these banks a comparative advantage in terms of costs or skills. This contrasts with “diversified banks,” which provide a broad range of different products and services to take advantage of economies of scale and scope that result from the mixing of different activities. Based on these two dimensions, banks can broadly be categorized into four different groups.

The first group comprises “specialized retail banks.” These banks offer a limited number of products to retail customers. Examples of specialized retail banks include auto banks and consumer finance banks. These banks use the internet or business partners, such as car dealers or retailers, as their distribution channels, which allows them to operate without (or with only a small number of) branches and employees. Due to their limited local presence, they do not form close customer relationships, which would give them access to soft information. They use hard information from credit scores instead and engage in transaction lending with standardized products that allow them to generate economies of scale. This contrasts with “diversified retail banks,” which have a larger number of branches and are more closely located to their customers. This allows them to build up a close relationship with their customers, to gather soft information over time, which goes beyond publicly available information, and to provide a broader and more differentiated range of products and services. Diversified retail banks are, therefore, more likely to be relationship lenders. A characteristic that both of these types of retail banks share is that lending is usually their most important activity and customer deposits their most important source of funding. Because of their focus on lending and deposit-taking, they usually generate a large share of their operating income from interest income, while other sources of income are less important.

Investment-oriented banks may also be specialized or diversified. An example of “specialized investment-oriented banks” are investment boutiques. They usually focus on one or a small number of areas of investment products or on a specific group of wholesale customers. Depending on their specialization, they offer a variety of services. For example, while some banks may act as investment advisors, others may specialize in trading in certain assets or commodities. This contrasts with “diversified investment-oriented banks,” which offer a broad range of products that includes underwriting, mergers and acquisitions (M&A), merchant and corporate banking, investment management and trading with different types of wholesale

<table>
<thead>
<tr>
<th>Retail-oriented</th>
<th>Investment-oriented</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specialized</strong></td>
<td></td>
</tr>
<tr>
<td>Focus on a specific group of retail customers (e.g., consumers, car buyers and small enterprises)</td>
<td>Focus on one or a small number of investment products or on a specific group of wholesale customers</td>
</tr>
<tr>
<td>Typically provide standardized retail products (e.g., consumer credit and car loans)</td>
<td>Non-interest income much more important than for retail-oriented banks</td>
</tr>
<tr>
<td>Interest income is the most important source of income (non-interest income often unimportant)</td>
<td>Primarily wholesale-funded (customer deposits unimportant)</td>
</tr>
<tr>
<td>Customer deposits are the primary source of funding (wholesale funding unimportant)</td>
<td></td>
</tr>
<tr>
<td><strong>Diversified</strong></td>
<td></td>
</tr>
<tr>
<td>Provision of a broader range of differentiated retail products and services (e.g., loans, mortgages and financial advisory) to different types of retail customers</td>
<td>Provision of a broad range of products that includes underwriting, M&amp;A, merchant and corporate banking, investment management and trading to different types of wholesale customers</td>
</tr>
<tr>
<td>Interest income is the most important source of income; non-interest income usually more important than for specialized retail banks; more diversified income structure</td>
<td>Non-interest income much more important than for retail-oriented banks (more diversified income structure than specialized investment-oriented banks)</td>
</tr>
<tr>
<td>More diversified funding structure, but customer deposits still most important source of funding</td>
<td>More diversified funding structure, but wholesale funding still the most important source of funding</td>
</tr>
</tbody>
</table>

Table 1: Systematization of business models in banking
Source: The table systematizes banks’ business model based on their products and the customers served (retail- versus investment-oriented banks) and the degree of specialization (specialized versus diversified).
Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

Because of the services provided, investment-oriented banks generate a significantly larger share of their operating income from non-traditional activities than retail-oriented banks. They are also more dependent on non-deposit funding and more oriented toward transaction banking (as opposed to relationship banking).

The business model matrix presented in Table 1, of course, only gives a simplified picture of the business models that exist in the banking sector and their main characteristics. In practice, business models are much more complex and differ along several other dimensions. The decision to engage in transaction or relationship banking, for example, is usually not that clearly defined, and the decision on how much of each the bank does depends on factors such as technology, regulation and competition (Boot (2000)). Boot (2000) finds that banks are more likely to be relationship oriented if they face competition from other banks, while competition from capital markets leads to a greater transaction orientation. Consequently, even though banks may mix relationship and transaction banking, investment-oriented banks tend to be more active in transaction banking, while relationship banking is more important for retail banks.

The decision to engage in retail or investment banking is often not that clear-cut either, and banks often mix both types of activities. Depending on the importance of these activities, we call these banks either “retail-oriented universal banks” or “investment-oriented universal banks.” Universal banking is the dominant business model in Europe. This contrasts with the U.S. where for a long time the Glass-Steagall Act prevented the emergence of

<table>
<thead>
<tr>
<th></th>
<th>All banks</th>
<th>Large banks</th>
<th>Medium-sized banks</th>
<th>Small banks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of banks</td>
<td>3,691</td>
<td>37</td>
<td>513</td>
<td>3,141</td>
</tr>
<tr>
<td>Market share of banks (in % of total assets)</td>
<td>100.00</td>
<td>74.32</td>
<td>22.58</td>
<td>3.10</td>
</tr>
<tr>
<td>Market share of banks (in % total number of banks)</td>
<td>100.00</td>
<td>1.00</td>
<td>13.90</td>
<td>85.10</td>
</tr>
<tr>
<td>Average banks size (in € billions)</td>
<td>9.73</td>
<td>723.80</td>
<td>15.67</td>
<td>0.34</td>
</tr>
</tbody>
</table>

**Assets (in % of total assets)**

- Cash and cash balances with central banks: 3.30, 3.76, 2.10, 0.76
- Available-for-sale assets: 4.75, 4.12, 7.44, 5.49
- Financial assets held for trading: 21.43, 25.45, 4.83, 1.24
- Total loans and advances: 55.86, 51.38, 68.95, 69.41
- Held-to-maturity investments: 0.82, 0.33, 2.83, 4.55

**Liabilities (in % of total assets)**

- Total deposits from credit institutions: 9.33, 8.15, 12.62, 13.87
- Total deposits other than from credit institutions: 39.43, 35.83, 47.20, 70.62
- Total equity: 5.0, 4.77, 5.39, 7.59

**Income structure and net interest margin**

1) Operating income comprises net interest income, net fee and commission income and net trading and foreign exchange income. The net interest margin is net interest income as a percent of total assets.

2) It should be noted that the indicators of loan quality might be driven by differences between national definitions of impaired assets (non-performing and doubtful assets) and provisions.

**Indicators of loan quality**

- Gross total doubtful and non-performing loans (as % of total debt instruments and total loans and advances): 3.92, 3.70, 4.65, 2.79
- Total loss provisions (as % of total gross doubtful and non-performing loans): 50.38, 48.73, 54.31, 43.88

Table 2: Descriptive statistics (2011)


The table shows descriptive statistics for several bank characteristics in 2011. Banks are broken down into three different groups by size. The first group of banks consists of small banks with assets of less than €1b. A second group of banks comprises medium-sized banks with assets ranging from €1b to €100b. The third group consists of very large banks with assets of more than €100b.
of universal banks and led to the separation of investment and commercial banks, with the latter providing retail and wholesale lending for both households and large corporations. The adoption in 1999 of the Gramm–Leach–Bliley Act did away with the traditional separation between investment and commercial banks in the U.S., allowing for the establishment of universal banks that combined a wide range of banking activities in one bank holding company. Critics argue that this has allowed U.S. investment banks to use insured deposits to fund their trading activities, which many believe to be a major cause for the 2007/2008 financial crisis. To prevent this cross-subsidization, the Dodd–Frank Act was enacted in 2011. An important part of this Act is the Volcker Rule, which prohibits insured depository institutions and their affiliates from engaging in proprietary trading. Even though the Volcker Rule does not go as far as the Glass–Steagall Act in separating commercial and investment banking, it will significantly limit banks’ activities.

4. Descriptive analysis of business models in the E.U. banking sector

In this section, we present some indicators that are commonly used to describe banks' business models. For this purpose, we use data from the ECB Consolidated Banking Statistics (ECB (2013)), which split banks into three groups by size. The first group of banks comprises small banks with assets of less than €1b. Typical examples of small credit institutions are savings and cooperative banks, which mostly operate in individual regions of a country. A second group of banks comprises medium-sized banks with assets ranging from €1b to €100b. In contrast to many small banks, these banks often operate nationwide. The third group consists of very large, internationally-operating banks with assets of more than €100b. We will shortly see that small banks are more active in retail banking, while medium-sized and large banks have a greater focus on investment banking activities.

First, however, we want to give a brief overview of the relative importance of small, medium-sized and large banks. Table 2 demonstrates that a large proportion of banks in the E.U. are small banks. Based on their assets, they are unimportant, however. For example, even though 85% of all credit institutions in the euro area are small, they only account for 3% of the assets. This, however, does not imply that small banks should be ignored in the financial stability assessment. Even though the failure of an individual small bank is less likely to have systemic implications, small banks’ distress might become systemic if a large number of them fail at the same time owing to exposures to certain asset classes, they have systemic connections to other intermediaries or are not immediately replaceable (Acharya and Yorulmazer (2007)).

Table 2 points to further significant differences between small, medium-sized and large banks. First, small and medium-sized banks have a much higher loans-to-assets and deposits-to-assets ratio than large banks. This is evident in both cases of banks that focus on retail banking. The positive correlation between lending and deposit-taking activities is consistent with theoretical models that predict that banks engaged in relationship lending can enhance their stability by combining lending and deposit-taking activities (Song and Thakor (2007), and Kashyap et al. (2002)). Large banks, in contrast, are much more active in investment banking and trading activities, which is reflected by their holdings having significantly higher proportions of trading assets. The share of assets held to maturity, in contrast, is significantly lower. Large banks also hold a significantly higher proportion of cash and cash balances with central banks and available-for-sale assets than small banks, which ensures the generation of liquidity within a short period of time at a predictable value. This is particularly important for banks that obtain funding primarily on the wholesale markets, since wholesale funds are more likely to be withdrawn prematurely and are less stable than deposits, not the least because they are protected by deposit insurance (Shleifer and Vishny (2010)). Consistent with that, we find that large banks have a much higher ratio of bank deposits to total assets than small and medium-sized banks.

In summary, business models differ considerably with bank size. While large banks are more active in non-traditional activities, such as trading and investment banking, and fund their activities on wholesale markets, small and medium-sized banks concentrate on lending activities and use deposits as their primary source of funding. Their retail orientation is also demonstrated by the share of net interest income to total income.

---

6 Trading assets are assets that have been acquired by the bank with the purpose of resale in the near term in order to profit from short-term price movements. They include bonds and other fixed-income securities, shares, and other variable-yield securities held for trading purposes.
Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

Operating income, which is significantly higher for small banks. Large banks, on the other hand, have a large non-interest income share, which is characteristic for investment-oriented banks. Non-interest income includes activities, such as income from trading and securitization, investment banking and advisory fees, brokerage commissions, venture capital and fiduciary income, and gains on non-hedging derivatives. With respect to the matrix of business models presented above, large banks, therefore, have a more investment-oriented business model, whereas small banks are more active in retail banking.

It should be said, however, that there may also be small banks that provide specialized investment banking services, and retail banks that are quite large in size. On average, however, bank size is a good proxy for distinguishing between retail-oriented and investment-oriented banks, since many characteristics describing these business models are related to their size.7

4.1 Impact of different business models on risk and return in the E.U. banking sector

The previous analysis indicates that business models differ significantly with bank size. To find out which banks have the greatest need to change their business models, we now analyze how banks performed during the financial crisis. We do so, first, by looking at indicators commonly used to assess bank performance. In the second step, we present recent findings in the academic literature that analyzes the impact of business models on bank risk and return.

An indicator that is often used to measure bank performance is pre-tax return-on-equity (RoE). Figure 2 suggests that large banks performed significantly worse than small and medium-sized banks in 2008, but also showed higher returns in 2009 and the following years. This contrasts with medium-sized banks, which performed better in 2008, but whose RoE decreased significantly thereafter. Small banks reported a negative RoE in 2011 as well. However, compared to large and medium-sized banks their RoE was much more stable over time, indicating that small banks were much less exposed to income fluctuations than large and medium-sized banks. Importantly, the better performance of large banks is driven not only by greater leverage, but also by a higher return-on-assets (RoA) than small and medium-sized banks.

This comparison of RoE suggests that large banks were particularly hit by the financial crisis in 2008. As markets stabilized and stock prices recovered, the profits of large banks rebounded and exceeded those of small and medium-sized banks. Their profits started to decline as of 2009 because of the economic downturn that increased credit risks and led to loan losses. It was particularly the medium-sized banks that saw loan losses, as indicated by the higher ratio of doubtful and non-performing loans and loss provisions in Table 2.8 The return of small banks, in contrast, was much more stable. Overall, the descriptive analysis suggests that large banks have more volatile returns, and therefore, seem to be more risky than small banks. Small banks, in contrast, have more stable returns, but are also

7 In the empirical banking literature, bank size is often used as a proxy for a bank’s ability to process soft information. Berger et al. (2005a, b), for example, show that smaller firms borrow from smaller banks and that while smaller banks do increasingly use credit scoring methods, they still have stronger relationships with their borrowers than large banks. Larger banks are less involved in relationship banking, because they are less capable of passing along soft information within the hierarchy (Berger and Black (2011)). In contrast to small banks, they focus more heavily on hard facts, such as credit ratings, and are engaged in transaction banking (Brown et al. (2012)).

8 Please note that these results should be interpreted with caution, since they might be driven by differences between national definitions of impaired assets (non-performing and doubtful assets) and provisions.
less profitable than large banks, which is consistent with the view that retail banking is a relatively stable, but also low-return, activity [Hirtle and Stiroh (2007)]. It should be noted that there were also small banks that reported large losses during the financial crisis. These banks were mostly specialized in mortgages and considerably increased their lending in the pre-crisis period. Furthermore, not all large banks performed badly, relatively. Some large and more retail-oriented banks were also less affected by the crisis. This indicates that it is not necessarily the size that matters for bank risk, but rather the extent to which banks diversify their activities.

4.2 Literature on income diversification in banking
The previous section indicated that levels of diversification might have an important influence on bank risk and return. To further investigate this hypothesis, we now take a closer look at the academic literature on diversification in banking. This literature usually examines the impact of geographic and income diversification on bank risk and return. In this paper, we focus on the literature on income diversification. This literature analyzes whether the expansion into non-traditional activities has made banks more stable and allowed them to generate higher returns.

Theoretically, the benefits of income diversification are appealing. Owing to a better diversified income structure, banks are able to recoup losses in some business areas through alternative sources of revenue in other areas. For example, by expanding into investment banking, banks can offset declining revenues in retail banking. This should lower the volatility of returns and lead to higher risk-adjusted returns. In theory, therefore, diversification into non-interest income activities, such as investment banking and trading, is beneficial for banks. In practice, however, the benefits of income diversification are less clear. Several authors argue that non-interest income is usually more volatile than interest income and allows banks to be more leveraged, because regulators often require banks to hold less capital against non-interest income activities [DeYoung and Roland (2001)].

Overall, therefore, it is an empirical question whether expanding into non-interest income activities makes banks more stable. For this reason, many empirical studies have been published over the past few decades. Most of these studies find only little evidence of gains from diversifying into non-interest income. Motivated by the financial crisis, more recent studies examine not only the impact of income diversification, but also how the diversification of funding sources affects banks’ risk and return. It is not clear whether diversification into wholesale funding makes a bank more or less risky, either. Experience from the financial crisis suggests that banks that are dependent on deposits for funding are stable, because customer deposits are less likely to be withdrawn prematurely than wholesale funds. The price of wholesale funds also adjusts more quickly to reflect a bank’s riskiness. Customer deposits, in contrast, are more stable and slower to be repriced, not least because they are protected by deposit insurance [Shleifer and Vishny (2010)] and held for liquidity purposes [Song and Thakor (2007)]. However, according to theory, wholesale funding may also reduce risk-taking if sophisticated wholesale financiers are better at monitoring than small depositors [Calomiris and Kahn (1991)]. Similar to the results for income diversification, the academic literature provides mixed evidence on the effects of funding diversification on bank risk and return as well.

Altunbas et al. (2011), for example, find that banks that were more dependent on non-deposit funding were more likely to fail during the financial crisis, while a better diversification of income sources was found to increase their stability. Demirgüç-Kunt and Huizinga (2010) find some risk diversification benefits at very low levels of non-interest income and non-deposit funding as well. For most banks, however, a higher share of non-interest income and more deposits as funding is beneficial. This is consistent with the findings of García–Herrero and Vazquez (2013) that smaller banks that were less profitable than large banks, which is consistent with the view that retail banking is a relatively stable, but also low-return, activity [Hirtle and Stiroh (2007)]. It should be noted that there were also small banks that reported large losses during the financial crisis. These banks were mostly specialized in mortgages and considerably increased their lending in the pre-crisis period. Furthermore, not all large banks performed badly, relatively. Some large and more retail-oriented banks were also less affected by the crisis. This indicates that it is not necessarily the size that matters for bank risk, but rather the extent to which banks diversify their activities.

9 These banks are mainly located in E.U. countries that experienced real-estate bubble in the pre-crisis period, such as the Spanish cajas.
10 In Spain, for example, Banco Santander and Banco Bilbao Vizcaya Argentaria (BBVA), the country’s two largest banks, were much less affected by the crisis than the considerably smaller cajas. Ayadi et al. (2011), for example, show that among Europe’s largest banks those identified as retail banks performed better and were more stable during the financial crisis.
11 For papers on geographic diversification in banking see García–Herrero and Vazquez (2013) and Bilinighausen and Köhler (2012).

12 Many studies focus on U.S. banks [see DeYoung and Roland (2001), DeYoung and Rice (2004), Goodhart et al. (2008), Stiroh (2004a, b) and Stiroh and Rumble (2006)]. For Europe, the evidence is also mixed. Lepetit et al. (2008), for example, show that banks that have expanded their non-interest income activities are more risky than banks that mainly supply loans. Mercieca et al. (2007) obtain similar findings for a sample of small European banks. Chiorazzo et al. (2008), in contrast, find that Italian banks will have significantly higher risk-adjusted returns. For Germany, Busch and Kick (2009) show that savings and cooperative banks are significantly more profitable as well if they increase their share of non-interest income, while they find no impact of non-interest income on commercial banks’ profitability. There is also evidence to suggest that banks from developing countries benefit from increasing their non-interest income share [Sanya and Wolfe (2011)].
Köhler (2012) finds that diversifying into non-deposit funding has a different impact as well. Common to both studies is their focus on listed banks, which are usually larger and have a more investment-oriented business model. This might have an important implication on how non-interest income affects bank risk.

While investment-oriented banks might be better off increasing their share of interest income, unlisted banks such as savings and cooperative banks might benefit from expanding into non-interest income. Unlike investment-oriented banks, they generate most of their income from retail banking activities and mainly depend on interest income. For these banks, a higher share of non-interest income might be beneficial, since it would make them less dependent on interest income and on movements of interest rates and net interest margins. Expansion into non-interest income activities may also make these banks more resilient to overall economic conditions that affect their loan portfolios. This suggests that banks that focus on retail banking might be affected differently by an increase in non-interest income than investment-oriented banks, which already have a high share of non-interest income. Hence, to come to more general conclusions about the impact of non-interest income on risk, it is important to enlarge the sample of banks and to include listed and unlisted banks.

Köhler (2012) takes this as a starting point to analyze the effect of non-interest income on risk and return in the E.U. banking sector. Using a large sample of listed and unlisted E.U. banks, he shows that smaller banks, in general, and savings and cooperative banks, in particular, become significantly more stable if they increase their share of non-interest income.13 Investment banks, in contrast, become significantly more risky. They do not only report a significantly higher non-interest income share, but also significantly differ from retail-oriented banks in terms of their activities. For example, while retail-oriented banks usually earn account administration, loan and consultancy fees and commission income from the sale of insurance products, investment-oriented banks derive most of their non-interest income from underwriting, treasury management, securitization and clearing, and other transaction-related services. Income from the latter set of activities is usually more volatile, because it tends to be more cyclical and closely linked with market evolution. This indicates that the risk characteristics of these two sets of activities are fundamentally different.

Further evidence in support of this hypothesis is provided by DeYoung and Torna (2013) for U.S. banks. They test whether banks that have a higher share of income from non-traditional banking activities were more likely to fail during the financial crisis. They find that it is not the non-interest income per se that made banks more likely to fail, but rather the type of non-interest income. More specifically, they find that a higher share of income from asset-based non-traditional activities, such as investment banking and asset securitization, makes distressed banks significantly more likely to fail, while a higher share of fee-based non-traditional activities, such as insurance sales significantly reduced the probability of failure. The latter usually represent a large share of non-interest income generated by retail-oriented banks.

In summary, most empirical studies find only little evidence of gains from diversifying into non-interest income businesses. This is because the greater volatility of these different sources of income, combined with the greater complexity and leverage of banks that have a higher proportion of non-interest income will offset the benefits of diversifying into non-interest income businesses. This implies that a higher share of non-interest income should make banks more risky. A more recent paper that includes data for the financial crisis indicates that this particularly applies to larger and more investment-oriented banks. Smaller and more retail-oriented banks, in contrast, might become more stable if they diversify into non-interest income. It is important to note that this does not imply that retail banks should expand into investment-banking and other areas in which they have little experience or comparative advantage, since this result might in fact make them more, and not less, risky (Stiroh (2004a), Mercieca et al. (2007) and Goddard et al. (2008)). Instead, they should increase their share of fee-based non-interest income. The effect of income diversification on a bank’s risk and return, therefore, crucially depends on that bank’s overall business model.

---

13 Köhler (2012) finds that diversifying into non-deposit funding has a different impact as well. While retail-oriented banks become significantly less stable if they increase their share of non-deposit funding, investment banks become significantly more stable. These findings indicate that it is important to enlarge the sample of banks and to include different types of banks with different business models to come to general conclusions about the effects of non-interest income and non-deposit funding on bank stability.
5. Restructuring business models in the post-crisis period

Business models have significantly changed over the past decades, as banks have become increasingly more active in non-traditional activities. While this has allowed banks to offset declining net interest margins, it has also exposed them to greater income volatility. Driven by an increase in demand, rising stock prices and a low stock market volatility many investment-oriented banks increased their capacities in the pre-crisis period. Due to cyclical reasons, such as the weak economy, low interest rates and less trading activity, many of these banks now have to cut costs and reduce their staff to adjust their capacities to the post-crisis reduction in demand for their products. According to a recent survey by EY (2013), for example, almost 50% of all European banks are planning to cut jobs over the next six months. Reducing costs and cutting jobs alone may, however, not be sufficient. For structural reasons, investment-oriented banks need to rethink their business models in general.

An important structural reason is the implementation of the Basel III banking regulations, which will significantly raise capital requirements. This will make business more costly and reduce the extent to which banks can raise their profitability by increasing their leverage. Growth will also be limited by higher liquidity requirements, which reduce the extent to which banks can fund their activities by wholesale funds and make customer deposits a more important source of funding. Banks also have to rethink their business models, because the recommendations of the Liikanen Group (2012), if implemented, will force them to move their trading activities into a separate legal entity. This will make business more costly and reduce the extent to which banks can raise their profitability by increasing their leverage. Growth will also be limited by higher liquidity requirements, which reduce the extent to which banks can fund their activities by wholesale funds and make customer deposits a more important source of funding. Banks also have to rethink their business models, because the recommendations of the Liikanen Group (2012), if implemented, will force them to move their trading activities into a separate legal entity. This will make business more costly and reduce the extent to which banks can raise their profitability by increasing their leverage. Growth will also be limited by higher liquidity requirements, which reduce the extent to which banks can fund their activities by wholesale funds and make customer deposits a more important source of funding. Banks also have to rethink their business models, because the recommendations of the Liikanen Group (2012), if implemented, will force them to move their trading activities into a separate legal entity. This will particularly affect banks with a large trading exposure.  

Overall, the recent regulatory reforms will particularly affect large, investment-oriented banks, since they are more active in trading, more dependent on wholesale funding and more highly leveraged. This, however, does not imply that small, retail-oriented banks do not need to change their business models. Because the lending and deposit markets will likely become more competitive, these banks are under pressure to rethink their business models as well. The important difference between those types of banks, however, will lie in how they have to change their business models. Because of that, we will now separately outline possible strategies with which banks can restructure their business models in the post-crisis period.

5.1 Possible strategies for large banks

The greatest pressure to restructure their business models is currently on large banks. As argued above, these banks tend to be more active in investment-banking and trading. Due to pressure from regulators and investors, they will likely have to reduce their dependence on non-interest income, in general, and trading income, in particular. The results from the recent academic literature suggest that this should lead to a better diversification of income sources and make them more stable. Expansion into retail banking activities is, however, difficult for several reasons. First, as outlined at the beginning, net interest margins have decreased significantly over the past few decades. This trend is likely to continue. Second, in order to gain market share, banks have to offer more attractive interest rates than their competitors, which further eats into their net interest margin. Moreover, large banks will find it hard to build close relationships with their customers, because retail banking was long seen as less profitable than investment banking and, therefore, abandoned by many banks. As argued above, large banks are also at a disadvantage relative to small banks in processing soft information and building relationships with their customers. This means that they are more dependent on hard information to evaluate the creditworthiness of potential borrowers and are more focused on transactional banking, which is more competitive. This is reflected by their net interest margin, which is significantly lower than that of small banks (see Table 1). The move toward retail banking will, therefore, at least in the short- and medium-term, likely put additional pressure on the profitability of large banks.

Because of this outlook, large banks should not completely abandon investment banking activities, but rather add lending and deposit-taking to their business mix to better diversify their income structure. The challenge for large banks is thus to mix investment banking with traditional commercial banking activities while maintaining a sufficient level of profitability and being sufficiently diversified. One way to make profits and maintain a diversified portfolio of activities is to refocus on core activities and core markets in which they have comparative advantages.

---

14 The Liikanen Group recommends that if a bank’s assets held for trading and available for sale exceed a threshold of 15% to 25% or €100bn then the bank would be referred to a supervisor to decide whether it needs to ring-fence its trading activities from its deposit-taking side.
over their competitors in terms of their expertise and experience or size. This means that banks have to close unprofitable areas and sell business lines in which they are too small, while in core businesses in which size is critical for success (e.g., consumer lending, global custody, asset management and payments) acquisitions may help them to strengthen their comparative advantage. Overall, the restructuring of business models is a difficult and time-consuming process and is occurring at a time when overall economic and financial conditions are weak and bank profitability is already low.

5.2 Possible strategies for small banks

In contrast to large banks, small banks usually focus on retail banking. Because investment banking allows banks to offset declining net interest margins and to generate higher returns, their business model has long been regarded as old-fashioned. This view has changed over the past few years, since the financial crisis has revealed the weaknesses of the investment banking business model. This, however, does not mean that the business models of small banks do not need to be changed. In fact, because the structural trends in the banking sector outlined at the beginning of the article are likely to persist, competition in lending and deposit markets will increase further. In recent years, the increase in the level of competition has become even stronger, since larger and more investment-oriented banks have rediscovered the merits of retail banking and increased their lending and deposit-taking activities to better diversify and to meet regulatory requirements, such as the Liquidity Coverage Ratio (LCR) and Net Stable Funding Ratio (NSFR). However, higher liquidity requirements will not only increase competition, but will also significantly reduce the possibilities for small banks to engage in maturity transformation to prop up their net interest margin. This will put additional downward pressure on their net interest margin, which is currently already low due to the low levels of interest rates. Since it is not clear for how long interest rates will remain low, and because of structural developments in the banking sector, the net interest margin will likely be low for the foreseeable future. Probably, this will particularly affect small banks, since they generate most of their income from interest and

are, therefore, heavily dependent on movements of interest rates and net interest margins.

To ensure sufficient profitability in the future, small banks, therefore, have to restructure their business models as well. The academic literature suggests that one possible way for small banks to generate higher returns and to become more stable is to increase the share of non-interest income. In the past, mostly large banks have expanded into non-interest income activities. This might have been driven by the fact that large banks usually earn lower net interest margins than small banks, which due to their focus on relationship banking, have some degree of market power that might have deterred them from increasing their non-interest income share significantly (see also Table 1). However, given the changes in customer behavior and the growing importance of transactional banking, their comparative advantage over large banks will likely decrease, and their interest margin will shrink. A more diversified income stream is also beneficial because it makes them less dependent on the level of interest rates. This should also reduce their incentive to increase credit risks during periods of low interest rates in search of yield (Delis and Kourtas (2011)). Banks may also be less likely to engage in maturity transformation in order to offset declining interest margins, thus possibly reducing interest rate risk.

Overall, therefore, small banks have to restructure their business models in a manner different from that which is beneficial to large banks. The discussion on the inherent riskiness of investment banking activities and the reform proposals of the Liikanen Group (2012) that aim to reduce these risks should, therefore, not distract small banks from the fact that a higher share of non-interest income may help them better diversify their income and to become more stable. It is important to note that this does not mean that small banks should expand into investment banking, as many large banks did over the past decades. Owing to their small size and their lack of expertise and experience in investment banking, they are at a disadvantage compared to large banks in this business and might become less stable (Mercieca et al. (2007) and Goddard et al. (2008)). A better strategy for these banks to expand the share of their non-interest income would be to build on their close relationships with their customers and to cross-sell additional financial services demanded by retail customers.

---

15 In Germany, particularly cooperative and savings banks rely on income from maturity transformation. According to Memmel (2011), their median share of earnings from term transformation was in the range of 13% and 15% between 2005 and 2009, while it amounted to less than 5% for the median private commercial bank.
Cross-selling may be the most practicable way to increase profits in stagnating or shrinking markets, in which lending volumes can only be extended if interest rates are reduced below the level of their competitors or if loans are granted to borrowers that have not received a loan from other banks because they asked for loan rates that are too low or provided insufficient collateral relative to their credit quality. Cross-selling may thus represent a strategy with which banks can increase their profits without incurring greater risks from lending activities. Increasing the rate of cross-selling and the share of non-interest income may, however, not be enough to ensure sufficient long-term profitability. Small banks also have to become more cost-efficient. So far, they can compensate for their cost disadvantages over large banks through higher net interest margins. The market power of small banks will, however, likely decrease if customers become more price-sensitive and less loyal to their banks. Furthermore, although proximity to a bank is still important, other distribution channels, such as online or mobile banking, have become more important over time. Because of the variety of different distribution channels, nowadays it is the customer that decides how to interact and not the bank. Banks have to react to these changes and to rethink the role of their branches. It is likely that this will change their function and reduce the number and the size of their branches. For banks, these changes will open up opportunities to improve their cost efficiency. However for smaller banks, in particular, these changes might be challenging, since they need to cut branches and staff while maintaining their comparative advantage over large banks owing to their close customer relationships. The closing of branches and job cuts, furthermore, will not alleviate the increasing burden of non-branch-related and non-staff-related overhead costs that arise, for instance, from IT infrastructure. Technological outsourcing is one strategy to reduce such overhead and to allow banks to refocus on their core competencies. Outsourcing is, however, not risk-free, and the potential for outsourcing is limited, since critical and complex functions cannot be outsourced. Furthermore, outsourcing always requires some sort of standardization, which reduces flexibility. Additional overhead costs arise as a result of the new Basel III regulations, since banks will probably need to recruit extra staff in order to ensure compliance with the new regulatory framework. These costs are usually disproportionally higher for small banks, since they will most likely not be able to completely pass the higher regulatory costs through to their customers (Santos and Elliott (2012)). For these reasons, a further consolidation will likely be necessary among small banks in future.

5.3 Possible impact on the structure of the E.U. banking sector
The restructuring process will not only affect banks' individual business models, but will also have an impact on the structure of the E.U. banking sector as a whole because it will likely lead to M&A. Because many banks are unloading their non-core businesses, the number of investment opportunities is expected to rise going forward. This will particularly affect large, investment-oriented banks, because they face the greatest pressure to become more streamlined and focused. They are also disproportionately affected by higher capital and liquidity requirements that force them to significantly deleverage and to reduce their capital usage.

While large, investment-oriented banks will, hence, most likely be the sellers, large, retail-oriented banks may be potential acquirers, since they have been relatively stable during the crisis. For example, the Spanish Banco Santander not only exited peripheral units by offloading its sub-scale businesses in Switzerland and the Czech Republic, but also strengthened its core markets by taking over the Polish Kredyt Bank from the KBC in 2012 to combine it with its Polish division. This indicates that some banks might also act as both seller and acquirer during their restructuring. Medium-sized banks with a solid capital base may also seek to strengthen their core businesses and markets by building scale in an effort to become more competitive.

Besides restructuring, the second major driver of M&As is consolidation. This will particularly affect small banks. Due to growing competition in lending and deposit markets, they are

16 For more information on the future of branch networks, see Köhler und Lang (2008).

17 According to ECB (2004), the greatest potential risk of outsourcing is the loss of control over the activities or services being outsourced and an undesirable dependency on the service provider.
under pressure to increase their cost efficiency as well. One strategy is to take over, or merge with, other institutions and to increase scale. A larger size may not only allow them to generate economies of scale, but also help smaller banks to better diversify [Demsetz and Strahan (1997) and Emmons (2004)]. The relationship between bank stability and size is not linear, however. As banks grow, the portfolio diversification effects of a larger bank size might be offset by diseconomies of scale due to managerial inefficiencies and greater complexity. Furthermore, as shown in the descriptive analysis, larger banks are more active in investment banking and trading and have a higher financial leverage, which might offset the diversification benefits of a larger bank size [Demsetz and Strahan (1997) and DeYoung and Roland (2001)].

In summary, M&A activity is likely to increase in the years to come. At present, however, it is still moderate, because many banks are unwilling to acquire businesses from other banks due to capital restrictions, uncertainty and investor skepticism. Furthermore, many banks may first wish to restructure their business model before they invest. The currently low level of M&A activity is reflected in Figure 3. The graph shows that M&A activity has significantly decreased since the beginning of the financial crisis, with the total deal value going down from almost €120b in 2008 to €12b in June 2013. Most targets that were acquired during this period were located in the U.K. (29%) and Spain (17%) and were crisis-driven, where national governments took control of failing banks. These transactions demonstrate that the crisis has changed the nature of M&A in the E.U. banking sector. While most deals in the pre-crisis period were strategic and aimed at building scale and expanding business lines, a large number of deals nowadays involve troubled banks that were (partly) nationalized through bailouts. In some deals, the government or regulatory authority did not directly nationalize the bank, but arranged or initiated a deal with private institutions. Government ownership will, however, likely be only a temporary phenomenon, since E.U. governments are expected to seek an orderly exit in the medium to long term. This should open up additional investment opportunities. In the future, the number of strategic deals should, therefore, rise again as many banks aim to restructure their business models and regain their strength to act as potential buyers on M&A markets.

Consolidation should also accelerate. In particular, in countries with a large number of banks relative to their population, a considerable potential for consolidation remains unexploited. According to Figure 4, these countries are primarily Austria, Denmark, Finland, Germany and Sweden, since the small numbers for Luxembourg and Ireland reflect their position as international financial centers. The need for further consolidation is also reflected by the small average bank size in these countries. In Austria and Germany, this is due to the large number of local savings and cooperative banks. According to EY (2013), most of the deals in these countries are expected to take place in more than three years from now, while most deals involving British and Spanish banks are crisis-related and expected to take place over the next 12 months.

To conclude, the financial crisis is a catalyst to M&A and will lead to a further consolidation of the E.U. banking sector. Small and large banks are differently affected by this process. Large banks will concentrate on their core activities and be left with less complex business models than pre-crisis. Because of comparative advantages, they will tend to be focused on corporate and investment banking. However, in order to become more diversified and to meet regulatory requirements, they are likely to increase their retail banking activities and sell basic, highly standardized financial products that allow them to...
generate economies of scale from their size. This transactional banking approach contrasts with small banks which, even though they tend to grow in size, will remain focused on more personalized, high-value-added products, such as small business loans and personalized investment and trust services, with profit margins driven by the willingness of customers to pay higher prices for these services. Relationship banking thus provides a niche for small banks that many large banks find less attractive or are less capable of providing (Berger and Udell (1995, 2002)). However, it is likely that the successful implementation of this strategy will require them to complement their traditional services with a wider scope of financial services that allows them to earn non-interest income and to partly offset the decline in net interest margins. Consequently, even though the crisis will accelerate the consolidation process significantly, the overall structure of the E.U. banking sector with a small number of larger, investment-oriented banks, on the one hand, and a large number of smaller, retail-oriented banks, on the other hand, will likely persist in future.

6. Conclusions
In this paper, we showed how business models evolved in the E.U. banking sector and how they need to be restructured to ensure sufficient profitability. At present, it is particularly large, investment-oriented banks that are under pressure to change their business models, since many business areas that guaranteed high returns pre-crisis have become unprofitable. We showed that these banks have to refocus on the core competencies and core markets in which they have comparative advantages. While maintaining their focus on corporate and investment banking, they will also need to be more active in retail banking to better diversify their income structure and to meet regulatory requirements. This move will not be an easy one and will likely put additional pressure on the profitability of large banks in the short and medium term.

Smaller, more retail-oriented banks have to rethink their business models as well, although they were much less affected by the financial crisis of 2007/2008. To become more profitable and stable, they should increase their share of non-interest income by cross-selling additional services or products to their customers. Diversifying into non-interest income is also beneficial because competition in retail banking will continue to be high for structural reasons even if interest rates start to go back up. It should be noted, however, that this does not imply that retail banks should expand into investment banking and other areas in which they have little experience or comparative advantage, since this might make them more, not less, risky. Small, retail-oriented banks and large, investment-oriented banks, therefore, differ in how they have to restructure their business models. Overall,

Figure 4: Consolidation in the E.U. banking sector (2012)
Source: ECB (2013). Figure shows the population per credit institution and the average bank size in the E.U.-15 banking sector in 2012. The yellow color bars represent the E.U.-15 average.
Business models in banking – how did they evolve and how do they need to be changed in the post-crisis period?

this implies that better-diversified banks are able to generate higher risk-adjusted returns and are more stable. However, while this may reduce idiosyncratic risk, it may also increase the level of systemic risk in the banking sector if banks diversify their activities in a similar way. This suggests that, from a financial stability perspective, it is necessary to have both diversified as well as specialized banks.

References

Berger, A., W. Frame, and N. Miller, 2005b, “Credit scoring and the availability, price, and risk of small business credit,” Journal of Money, Credit and Banking 37, 191-222
Demsetz, R. S., and P. E. Strahan, 1997, “Diversification, size, and risk at bank holding companies,” Journal of Money Credit and Banking 29(3), 300-313
ECB, 2013, ECB consolidated banking statistics
EY, 2013, European banking barometer.
Liikanen Group, 2012, High-level expert group on reforming the structure of the EU banking sector
Stiroh, K. J., 2004b, “Diversification in banking: is non-interest income the answer?” Journal of Money, Credit, and Banking 36, 853-882
Zephyr, 2013, “Database on M&A,” Bureau van Dijk
Valuation effects of termination of cross-listings

Roland Füss  
Full Professor of Real Estate Finance, University of St.Gallen

Ulrich Hommel  
Full Professor of Finance, Department of Finance, Accounting and Real Estate, EBS Business School

Jan-Carl Plagge  
Ph.D. Candidate, Department of Finance, Accounting and Real Estate, EBS Business School

Abstract
In the past, segmentation of capital markets incentivized numerous companies to cross-list in more integrated markets by promising better investor recognition and thereby lowering cost of capital and increasing stock liquidity. In recent years, however, more and more companies have decided to terminate these dual listings. In contrast to previous studies, which mainly focus on the benefits and/or drawbacks of the initial cross-listings, this paper investigates share price effects linked to the termination of cross-listings. A special focus lies on the characteristics of companies' home markets to explain observed price effects. The empirical results show negative abnormal returns on the event date as well as negative cumulative abnormal returns in the weeks following the announcement of the termination of the secondary listing. The absolute value of the price reaction varies positively with foreign investors' difficulty to enter the companies' local market and the fraction of liquidity allocated to the foreign listing prior to the delisting event.
1. Introduction

Empirical studies on initial cross-listings have found positive abnormal stock returns around the announcement date of the secondary listing. This observation has typically been explained by the ability of these firms to overcome the market segmentation disadvantages of purely locally listed companies. Without the ability to tap international capital markets, companies resident in segmented markets are assumed to face higher equity costs imposed by local investors that cannot diversify unsystematic country risk. As companies are moving from a segmented to an integrated market, a positive revaluation effect sets in as the minimum required return of equity investors falls.

Given this reasoning, it is surprising that an increasing number of companies — not only well-known European corporations, such as Ahold, Air France, Bayer, British Airways, Danone and Fiat (Dobbs and Goedhart (2008)), but also companies resident in emerging markets — have in recent years decided to delist from well integrated stock exchanges, such as the New York Stock Exchange. Several factors may explain why the termination of a foreign listing should go hand in hand with a negative share price reaction. Predominant, seems to be a rise in the company's cost of capital, especially when the home market is not integrated at the time of the delisting. While potential implications of initial cross-listings have been extensively studied in the literature, the effects of delistings have been investigated by merely a handful of recent studies and with mixed results. This is mainly due to the fact that the trend to delist has only gained momentum in the last few years.

In this paper, we examine whether the termination of a secondary market listing leads to a reversal of positive share price reactions commonly observed around the initial cross-listing announcement. To our knowledge, this is the first study that explicitly analyzes the impact of capital market segmentation/integration on abnormal returns at the time of delisting announcements. This study in particular investigates the role that the degree of development and accessibility of local stock markets at the time of the delisting announcement have in explaining delisting-related abnormal returns. The expectation should be that abnormal returns are smaller, the higher the degree of the local market's segmentation at which the company terminates its cross-listing.

The empirical results show significant negative abnormal returns on the event date as well as negative cumulative abnormal returns in the weeks following the announcement of the termination of a secondary listing. The magnitude of the price reaction varies positively with the difficulty of foreign investors entering the companies' local market as well as the fraction of liquidity allocated to the foreign listing prior to the delisting event.

The remainder of the paper is structured as follows: The next section presents the economic rationale underpinning cross-listing and cross-delisting. Section 3 provides the determinants of price reactions observed around cross-listings and cross-delistings based on a review of the literature, and derives the corresponding hypotheses to be tested. Section 4 describes the data and the estimation approach. The empirical results are summarized and interpreted in section 5. Section 6 concludes.

2. Reasons to list and delist depository receipts

Related research has identified market segmentation as the primary motivation for companies to cross-list their shares. Companies are able to lower their cost of capital by making their stock tradable on exchanges characterized by a higher degree of financial market integration. Jorion and Schwartz (1986) define market integration as a state where investors earn the same risk-adjusted expected return from similar financial instruments, independent of the exchange on which the shares are traded. Similarly, Bekaert and Harvey (1995) argue that companies listed locally in completely integrated capital markets should have the same stock returns for a given level of risk irrespective of their domicile country. In this case, return variance is replaced by the stock's covariance to the integrated world capital market. In contrast, if a market is completely segmented, the stock's covariance with a given world factor does not exist and thus, investors’ expected returns are related to the company's total risk.

The reasons for market segmentation are manifold. They might arise from investors’ limitations, such as the lack of information about the target market, fear of expropriation, or discriminatory tax treatment abroad. Moreover, capital controls may restrict foreigners’ access to local capital markets and reduce their
freedom to repatriate capital and dividends. Some countries, predominantly emerging markets, also formally restrict the fraction of the local firm’s equity that can maximally be owned by foreign investors. Given the above definition based on the law of one price, it follows that a corporation’s decision to list shares outside of a segmented or semi-integrated market has direct consequences for the investor’s required rate of return and in turn for the firm’s cost of capital.

Based on Markowitz’ (1952) portfolio selection theory, local investors resident in segmented capital markets require a premium for investing in a local company’s equity and not being optimally diversified. Companies may, therefore, be able to significantly reduce their cost of capital by escaping local segmentation and tapping global capital markets. Unlike locally restricted investors, global investors do not require a risk premium for the specific country risk. Due to the lower expected cost of capital, numerous empirical studies observed abnormal positive returns causing a general rise in the corporation’s share price around the announcement date of a cross-listing. Given this evidence, it is worthwhile investigating why increasing numbers of companies are reversing their initial decision to cross-list and terminate their foreign listing.

One explanation often cited is the balance between the so-called “bonding” and the “loss of competitiveness theory.” “By subjecting themselves to U.S. laws and institutions, the controlling shareholders of foreign firms credibly bond themselves to avoid some types of actions that might decrease the wealth of minority shareholders.” [Doidge et al. (2008)] Companies that are resident in countries with lower legal and institutional standards should especially benefit from a commitment to stricter disclosure requirements linked to a secondary listing on one of the major U.S. stock exchanges. Meeting the more extensive disclosure requirements, however, also induces higher costs and complexity, e.g., due to the Sarbanes Oxley Act (SOX) to which companies cross-listed on U.S. stock exchanges have been subjected since 2002. This Act, enforced by the Securities and Exchange Commission (SEC), requires a strict internal control system. Compliance demands costly monitoring of internal business processes, which may result in a loss of competitiveness. Thus, the decision to terminate a secondary listing can indeed be legitimated when the net benefits related to maintaining a cross-listing are no longer present.

In particular, for companies locally listed in well integrated and easily accessible markets, there should be no need to overcome market segmentation by cross-listing their shares on a foreign exchange. Moreover, the standard of information disclosure required by exchanges in developed markets is already significant and companies typically disclose a high level of information to outside investors even without being SOX compliant. Thus, with the positive effects derived from increased integration and better disclosure being negligible, the net benefits of a cross-listing in the U.S. may easily become negative for companies locally listed on well-integrated stock exchanges. This may have driven larger firms that are listed locally in well-developed and integrated capital markets to delist in recent years.

3. Valuation effects around cross-delistings
3.1 From cross-listing to cross-delisting
The positive revaluation effect of an international listing induced various researchers to pay special attention to potential price reactions around the announcement date of the initial cross-listing. Torabzadeh et al. (1992), Lau et al. (1994), Miller (1999), Foerster and Karoly (1999), Litvak (2008), and Roosenboom and van Dijk (2009), for instance, observe positive abnormal returns around the announcement date of a depository receipt (DR) listing. Kadlec and McConnell (1994) even derive abnormal returns of about 5% in response to the cross-listing announcement.

This assumption is backed by Zingales (2006) who observed a significant decrease in the U.S. share of global IPOs between 2000 and 2005. Next to an increase in the competitiveness of non-U.S. capital markets, he relates his findings to an increase in the compliance costs for companies publically traded in the U.S.

4 The extremely difficult and often costly process of deregistering from the SEC has previously prevented companies from delisting. However, the trend to delist increased significantly in the wake of the “first significant deregulation of U.S. disclosure requirements since the passage of the 1933/1934 Exchange and Securities Acts: the 2007 Securities and Exchange Commission (SEC) Rule 12h-6. The Rule 12h-6 has facilitated foreign firms to deregister with the SEC and thereby terminate their U.S. disclosure obligations. [Fernandes et al. (2010)] Companies that formerly cross-listed shares on U.S. stock exchanges are since then allowed to “deregister with the U.S. Securities and Exchange Commission if less than 5% of global trading takes place on U.S. stock exchanges.” [Dobs and Goedhart (2008)]
In an attempt to disentangle the sources of abnormal returns around cross-listings, Miller (1999) finds that the extent of the stock price reaction depends on the stage of the local stock market development. He discovers positive revaluation effects to be significantly larger for firms listed locally in emerging markets as compared to those located in developed markets. Serra (1999) compares the revaluation effects of emerging market companies with a control sample of European firms that both cross-listed on U.S. and U.K. stock exchanges. In line with Miller (1999), she observes a significantly more pronounced effect of abnormal positive returns for dual-listed companies resident in emerging markets.

By focusing on a global set of firms cross-listing their shares on U.S. exchanges, Bris et al. (2007) attempted to determine the impact of market segmentation, liquidity, and controlling shareholders’ bonding on abnormal returns associated with cross-listings. Corroborating the findings of Miller (1999) and Serra (1999), they identify market segmentation to have the greatest explanatory powers with regards to abnormal returns. Doukas and Switzer (2000) investigate the effect of market segmentation along the time dimension by controlling for subsequent shifts in the degree of market integration. Assuming that positive revaluation effects decrease with increase in market integration, they observe price reactions related to the announcement of Canadian companies that simultaneously list their common shares on U.S. stock exchanges. However, since they do not find evidence for increasing market integration during the sample period 1985 through 1996, revaluation effects around the announcement of the secondary listings on a U.S. exchange are significantly positive.6

The observations made by Miller (1999), Sundaram and Logue (1996), Serra (1999), Bris et al. (2007), Litvak (2008) and Roosenboom and van Dijk (2009) all highlight the importance of the state of market integration in explaining the abnormal returns around the initial cross-listing. Consequently, it seems reasonable to assume that market segmentation should play a likewise important role in explaining abnormal returns around a cross-delisting decision.

Focusing on price effects around delisting announcements, Hostak et al. (2006) document positive abnormal returns for companies deregistering in the post-SOX period. By explicitly distinguishing between pre- and post-SOX periods, Marosi and Massoud (2008) observe less negative and less significant market reactions in the latter period. Hence, they conclude that the passage of the Sarbanes-Oxley Act has reduced the benefits of a U.S. listing and registration, particularly for smaller firms with lower trading volume and stronger insider control.

Doidge et al. (2008) show that de-registrations are generally associated with adverse stock price reactions, whereas “firms with better growth opportunities and larger financing deficits have significantly worse deregistration-related stock price reactions.” You et al. (2012) observe a negative price effect related to cross-delistings, which disappears again in the long run. Studying price reactions around the deregistration process, Fernandes et al. (2010) observe negative market reactions for companies located in countries with low disclosure requirements, as well as for firms from countries with civil law legal origin and with low levels of judicial efficiency. In contrast, stock price reactions are insignificant for firms located in countries with low investor protection.

3.2 Determinants of valuation effects around cross-delistings

As outlined in the previous section, several empirical studies have demonstrated that companies locally listed in emerging markets experience significantly higher positive abnormal returns as a result of their cross-listings on U.K. or U.S. exchanges as compared to firms locally listed in developed markets. It is assumed that this observation can be explained by emerging markets being less accessible for foreign investors and consequently more segmented than developed markets.

Applying this reasoning to cross-delistings, it is hypothesized that, analogous to share price reactions measured around the announcement of the initial cross-listing, abnormal returns observed in the process of a delisting are a function of the

5 To measure the degree of market integration, the authors applied a regime-switching model according to Bekaert and Harvey (1995).

6 Expanding the focus beyond the effects directly associated with the announcement of cross-listings, studies conducted by Alexander et al. (1988), Lau et al. (1994), Serra (1999), Foerster and Karoly (1999), Errunza and Miller (2000) and Sarkissian and Schill (2009) all find a significant decline in returns in the time-period following the listing. This observation supports the initially hypothesized reduction in the cost of capital for companies escaping from segmentation. While Foerster and Karoly (1999) observe an average loss in returns of 14% during the year following the cross-listing, Errunza and Miller (2000) even measure a decline of around 40% in the cost of capital after cross-listing.
accessibility of the firm’s local capital market. If investors want to stay invested in a company after it terminates its foreign listing, they need to transfer their investment to the company’s local exchange. The easier it is for foreign investors to transfer their investment to the local capital market, the more investors should be willing to do so in order to stay exposed to the company’s specific risk-return profile. This leads us to our first hypothesis:

**Hypothesis H1:** the less accessible a local capital market, the smaller (more negative) the abnormal returns related to the cross-delisting – the maturity of capital markets, as a major part of existing country classification systems, is a second dimension which should influence abnormal returns. Institutional investors often have internal requirements with respect to minimum liquidity thresholds that need to be fulfilled before an investment can be carried out. Thus, market liquidity can be interpreted as an indirect measure of market accessibility given that drops in turnover levels lead to the exclusion of certain investor groups.

**Hypothesis H2:** the less liquid the local capital market, the smaller (more negative) the abnormal returns related to the cross-delisting – the question of whether a foreign listing is beneficial for a given company also needs to be evaluated in light of the “bonding” and the “loss of competitiveness” hypotheses [Doidge et al. (2008)]. If local markets are already subjected to high regulatory standards, benefits from bonding should not exist, or should at least be less significant. The persisting costs associated with maintaining the foreign listing might therefore even lead to a negative net effect.

**Hypothesis H3:** the less developed the local market’s regulatory environment, the smaller (more negative) the abnormal returns related to the cross-delisting – the share of total liquidity aggregated over all listings of a company that relates to the ADR listing should play a decisive role in explaining abnormal returns given that it serves as a proxy of the cost of capital reduction related to falling country-specific risk premium. Related literature assumes that the information available for a given company is highest in its home market. Focusing on the liquidity allocation between local and foreign dual listings, Haling et al. (2008) hypothesize that the availability of information determines where a given stock is traded. Using the geographical distance between the two exchanges as a proxy for information density, they find liquidity in the foreign cross-listing to be the higher, the closer the

place of foreign trade to the local market of the company. Thus, to exploit this information-related advantage, investors should be more inclined to trade locally if both the local and the foreign exchange were similarly easy to access. In contrast, a flourishing foreign market should only emerge if the local exchange is relatively difficult to access. In this case, cross-delisting is likely to lead to an exit of foreign investors.

**Hypothesis H4:** the higher the fraction of depositary receipts (DR) liquidity, the smaller (more negative) the abnormal returns related to the cross-delisting – the literature typically ignores the fact that companies can have more than one foreign listing, which should have a dampening effect on the share price reaction following a cross-delisting. Take for instance a company locally listed in an emerging market with further foreign listings in developed markets. Terminating the DR program in the U.S. is, for instance, likely to shift liquidity to other developed markets, thereby reducing the magnitude of abnormal returns. What, therefore, matters is the fraction of total liquidity generated in emerging markets, as it reflects the difficulty of accessing a company’s stock after the termination of a foreign listing.

**Hypothesis H5:** the higher the fraction of total liquidity allocated to emerging market listings prior to the cross-delisting, the smaller (more negative) the abnormal returns related to the cross-delisting.

### 4. Data and methodology

To investigate the hypothesized effects, the study focuses on American Depository Receipts (ADR) formerly cross-listed on U.S. stock exchanges (Level II and Level III DR programs) as common vehicles for cross-listings. As a first step, we searched all active and inactive depositary receipts of companies domiciled outside the U.S. for delisting announcements over a 13-year period from 2000 to 2012. This leads to 457 DR observations, which either still exist or have been delisted in the past. Of those, 124 companies have been excluded as the delisting event was related to an acquisition or a merger event implying a dilution of the

---

7 Depository Receipts traded in the U.S. can either be traded as Level I, Level II or Level III DRs. While Level I DRs are traded over the counter (OTC), Level II and Level III DRs are listed on the NYSE or the NASDAQ. Level II DRs refer to existing shares locally listed in the cross-listed companies’ home market. “New DRs are created from deposits of ordinary shares in the issuer’s home market. Level III DRs are public offerings of new shares into the U.S. market.” [Depository Receipts – Reference Guide, JPMorgan (11)]
Valuation effects of termination of cross-listings

price reaction. Similarly, we discard another 34 companies as the delisting announcement was released together with other information, such as periodic financial results or corporate reorganizations. For 97 delisted companies, no announcements were available, and another eight companies were removed as their delisting coincided with insolvency. Overall, this has led to a sample of 194 delistings with historical price data actually being available for 147 of those companies.

Of the final selection, 121 delistings were purely voluntarily, whereas 26 delistings were a consequence of a violation of listing requirements imposed by the respective foreign exchange. Examples of such violations are the failure to file quarterly financial reports as required by the SEC or not meeting minimum stock price or liquidity requirements. 84 companies of the final sample were cross-listed on the NYSE and 63 on the NASDAQ stock exchange. The sample is geographically well diversified with shares locally listed in 26 different countries, of which 18 countries are classified as developed and eight as emerging markets according to the MSCI country classification standard.

To analyze valuation effects resulting from the termination of cross-listings we use an event-study approach following Lau et al. (1994), Miller (1999), as well as Roosenboom and van Dijk (2009). To assess abnormal returns around the cross-delisting event, the event-study methodology measures the deviation of the true company-specific and event-related return from an expected return undistorted by the specific delisting event. To estimate the expected return, a single-factor market model based on each listing’s local market-capitalization-weighted MSCI country index is constructed. Over the course of a pre-announcement date period (estimation period), each company’s returns are then regressed against the corresponding country index. The estimation period applied in this study ranges from $d_{-1000}$ to $d_{+1000}$ days (147 weeks) to $d_{-20}$ days (3.5 weeks). The corresponding event window ranges from $d_{-19}$ days (3.5 weeks) to $d_{+12}$ days (2.5 weeks). The regression coefficients are then used to calculate the expected return for the given stock. Abnormal returns (AR) are determined as: $AR_{id} = r_{id} - (a_i + \beta_{id} r_m)$ (equation 1), with $AR_{id}$ = abnormal return for stock $i$ on day $d$, $r_{id}$ = return on stock $i$ on day $d$, and $r_m$ = return of country $c$'s market index on day $d$. Various studies found that the announcement day $d_0$ rather than the initial listing date is the relevant event when it comes to analyzing abnormal returns associated with the initial cross-listing (see, e.g., Kadlec and McConnell, 1994; Miller, 1999; Doukas and Switzer, 2000; Roosenboom and van Dijk, 2009). It is assumed that the observed price effects associated with the overcoming of market segmentation will be incorporated in the company’s share price in the immediate proximity to the initial announcement. When it comes to studying the effect of delistings, the same reasoning should hold.

To test for the statistical significance, abnormal returns are averaged on the event week (day) $d_0$ over the respective sample of $N$ listings: $\overline{AR}_c = \Sigma_{i=1}^{N} AR_{id} / N$ (equation 2). As a robustness check, abnormal returns are standardized by dividing each listings abnormal return by its standard deviation determined over the estimation period $S(AR)$: $SAR_{i0} = AR_{i0} / S(AR)$ (equation 3). This approach scales abnormal returns by stock price volatility, which serves as a proxy measure for “confidence.” Consequently, scaled abnormal returns of stocks that display a high historical volatility are relatively lower compared to scaled abnormal returns of low-volatility stocks. Accordingly, the average standardized abnormal returns are defined as: $\overline{SAR}_c = \Sigma_{i=1}^{N} SAR_{i0} / N$ (equation 4).

---

8 Note that the simultaneous release of additional information might also affect stock prices and, therefore, does not allow us to properly assess the pure impact of the delisting announcement on prices of cross-listed shares. Excluding these cases is common practice in the related literature.

9 The MSCI country indices applied in the analysis are broad market indices covering 85% of the investable capitalization of the underlying market. Due to the broad market coverage, the event study should not suffer from selection bias.

10 The historical price data on the company level as applied in the analysis is taken from Thomson Reuters Datastream. The time series are adjusted for corporate actions such as stock splits.

11 The analysis focuses on weekly rather than daily abnormal returns as this allows for the measurement of price effects caused by market participants that had relevant information shortly before the official announcement was released (information leakage) or reacted with a short delay. This would, of course, also be captured by cumulative abnormal returns; however, cumulative abnormal returns cover a longer period around the announcement date which decreases their statistical significance due to the cumulating nature of standard errors. To assess the robustness of the model, the analysis is also conducted using daily returns. For the daily model, the estimation period ranges from $d_{-1500}$ to $d_{+11}$ days, whereas the event period ranges from $d_{-10}$ days to $d_{+10}$ days. Next to the models presented in the paper, the analysis was re-run based on a 60-day event period (daily returns) as well as based on a 10-week event period (weekly returns). Furthermore, the length of the estimation period has varied. The results remained qualitatively the same. To conserve space the results are not shown here, however, they are available from the Supplementary Internet Appendix available at http://www.sbf.unisg.ch/~media/Internet/Content/Dateien/InstituteUndCenters/SBF/Mitarbeitende.CV.etc/Effects_from_Cross_Listings_Fuus_Hommel_Plagge_Internet_Appendix.ashx.
To also capture effects that might either come from information leakage prior to the official announcement or from market participants that act over the course of several days enclosing the announcement date, the analysis is extended by taking into account cumulative abnormal returns (CAR). The cumulative abnormal returns are determined based on an aggregation of the average abnormal returns over M weeks (days) of the event window around the announcement date: $\text{CAR} = \sum_0^M \text{AR}_m$ (equation 5). To test for statistical significance, the following t-statistic is constructed: $t = \frac{\text{AR}}{S(\text{AR})}$ (equation 6).

Various approaches related to the determination of the standard deviation used to compose the test statistic are discussed in the literature. This paper applies four parametric methods to test the robustness of the results. First, the standard deviation is determined under the assumption of cross-sectional independence. Consequently, each stock’s volatility is determined based merely on its own historical observations. To also account for potential cross-sectional dependence among the average residuals, a second model is composed according to Brown and Warner (1985). The standard deviation of this second model is determined based on the average abnormal return across all returns and weeks of the estimation period rather than based on each single stock’s historical yield observations. As a third model, the aforementioned standardized abnormal returns are used. As their variance is by definition equal to 1, the standard deviation used to compose the test statistic is given by the inverse of the square root of the number of companies in the sample: $S(\text{AR}) = 1/\sqrt{N}$ (equation 7).

As pointed out by Serra (2002), the null hypothesis of zero abnormal returns is typically rejected too often in case the variance of stock returns increases during the event window. To overcome this problem, Boehmer et al. (1991) proposed to calculate the variance cross-sectionally within the event window rather than for each security over the time of the estimation period, which leads to our fourth model. To assess the hypotheses stated, we first divide the entire sample into two subsamples – companies locally listed in developed and companies locally listed in emerging markets. Abnormal returns are then calculated for each subsample individually. Potential differences between the abnormal returns of the two subsamples as a reaction to the cross-delisting provide a first indication on how market segmentation (H1) maturity (H2) and regulation (H3) of capital markets affect stock price movements as all three dimensions are part of conventional country classification systems. Compared to their developed counterparts, emerging markets are typically less integrated, display lower levels of liquidity and have an inferior regulatory system. While the comparison of abnormal returns based on subsamples provides the basis of an initial qualitative analysis, an Ordinary Least Squares regression model is constructed to explicitly test the above stated hypotheses for the entire sample.

To test the robustness of the regression results, abnormal returns are measured over different time periods surrounding the event date. The first regression model is composed based on cumulative weekly abnormal returns calculated over a 6-week period surrounding the event day, starting 3.5 weeks prior to and ending 2.5 weeks after the event date. A second regression model is based on cumulative abnormal returns over a 4-week period surrounding the event day, starting 3.5 weeks prior to the announcement date and ending with the announcement week. Finally, a third regression model is based on abnormal returns during the event week itself.

We include the following independent variables in the model specifications:

**Market Integration**: capital market integration is measured using three variables. As a first proxy for market integration, the Economist Intelligence Unit (EUI) criterion “access of foreigners to the local capital market” (AC) is applied. Secondly, the ease of capital repatriation (REP) from the Political Risk Services (PRS) Group is included. Capital repatriation should play a decisive role in assessing market access. The harder it is to repatriate

---

12 Formulas are based on Cowan (1992) and Serra (2002).
13 The Supplementary Internet Appendix available http://www.sbf.unisg.ch/~/media/Internet/Content/Daten/Institute/InCenters/SBF/Mitarbeitende_CV/etc/Effects_from_CrossListings_Fuess_Hommel_Plagge_Internet_Appendix.ashx provides a detailed overview of the different ways to determine the different standard deviations for hypothesis testing.
14 Countries are classified according to the MSCI country classification system. MSCI groups countries into three categories: developed, emerging and frontier markets. The allocation is done based on the criteria economic development, size and liquidity requirements, as well as market accessibility. We refer to the MSCI Market Classification Framework available under www.msci.com for a detailed breakdown of the respective criteria.
15 The inclusion of fitted values does not lead to a measurement error problem as the abnormal returns are included as the dependent variable. Consequently, the initial measurement errors are captured by the error term of this regression.
capital, the less willing foreign investors are to allocate money to that market given the uncertainty and difficulty of subsequent capital withdrawal. As third measure for market integration, the correlation (COR) of each country index with the world market (MSCI World index) measured over a three-year period based on weekly returns is used.

**Capital market liquidity:** capital market liquidity is measured as aggregated traded value scaled by GDP (MVOLL).

**Regulatory environment of the local market:** The regulatory environment of the local market is assessed using three criteria composed by the EIU. Investment protection (IP) evaluates to what extend property rights are in place in order to guarantee ownership. Insufficient investment protection should deter foreign investors from investing locally. Further, the transparency and fairness of the legal system (TFLS) is included to capture the reliability on the legal frame of a country. Lastly, the state of the financial regulatory system (FRS) is included as a direct measure for the effectiveness of a local investment.

**Liquidity allocation:** The fraction of DR liquidity (LSDR) is determined as DR turnover divided by the turnover of all listings issued by a given company aggregated over the six months preceding the delisting month. Similarly, the fraction of emerging market liquidity (LSEM) is the ratio of the sum of the turnover of all listings in emerging markets and the total turnover of all listings.

**Control variables:** A variety of variables controlling for factors that might also influence the price reaction related to the determination of a foreign listing is included in the regression. To control for firm characteristics, total assets (TA), the price-to-book ratio (PTB), returns on assets (RoA), as well as the leverage ratio (LEV) determined as total liabilities over total assets, are considered. The price-to-book ratio is added to control for growth opportunities. Companies that face more extensive growth opportunities should have a great need for external financing and should consequently rely more heavily on a broad investor base compared to companies with more limited growth opportunities. RoA is included to control for firm profitability. The termination of a foreign listing announced by a low-performing company can be interpreted as a negative signal concerning the company’s present state, and may consequently lead to a negative price reaction if investors decide to reduce their holdings. Similarly, the leverage ratio is included since the delisting announcement of highly indebted companies may send a negative signal about the company’s performance prospects to outside investors.

Roosenboom and van Dijk (2009) point out that the information disclosure requirements imposed by the SEC reduce information asymmetries, which might explain positive abnormal returns around the initial cross-listing. Thus, it can be assumed that a negative price reaction is caused by the announcement to terminate the registration with the SEC. To capture this effect, a control variable (SEC) is added, being 1 if the company announces to remain registered and 0 otherwise.

Stock price volatility (STDDEV) measured over a three-year window of weekly returns is included to control for the variability of price movements. Since volatility can be interpreted as a sign of greater uncertainty about a company’s financial outlook, the price reaction to a delisting announcement might imply positive or negative deviations from the fair value.

In reference to the competitiveness theory mentioned above, the costs associated with the delisting itself might also play a role in explaining abnormal returns. The introduction of Rule 12h-6 has significantly reduced the expense and time-to-implementation of a delisting. The resulting efficiency gains can be assumed to have a positive influence on abnormal returns. To control for this effect, a dummy variable (R12h-6) is included which equals 1 if the announcement date of the termination is prior to March 2007, the month R12h-6 went into effect, and 0 otherwise.

---

16 The Political Risk Services Group (PRS) as well as the Economist Intelligence Unit (EIU) offer country risk analytics widely used in academia as well as by political agencies.

17 For the purpose of comparison, the ordinal nature of the EIU variables has to be converted to binary variables, being 1 if the country receives the best possible score within the respective scaling and 0 otherwise.

18 A few companies do not explicitly mention their SEC deregistration in press releases when announcing their delisting. However, they typically refer to high administrative costs associated with the cross-listing as a reason to delist. Other companies are keen to mention that even after delisting, they will keep their SEC registration and accompanied disclosure requirements. Given this intention to signaling, it is assumed that companies that do not explicitly mention keeping their SEC status will terminate their registration.

19 Note that the effect of company-specific over- and under-reactions is partially already controlled for by the event-study model. Companies that tend to overreact relative to the market display a higher beta which generally raises the anticipated “normal” return around the event date and consequently decreases its abnormal return. While the market model only captures systematic risk, stock price volatility relates to a company’s total risk.
Focusing on Canadian companies cross-listed in the U.S., Mittoo (1997) observes an exchange-dependent liquidity reaction associated with the initial cross-listing. Given the importance of exchange-dependent characteristics and the fact that our sample contains two foreign exchanges, our model specification contains a dummy variable (EXCH) controlling for potential effects related to the former exchange (NYSE or NASDAQ). Further, a dummy variable (PLS) controlling for the post-listing status of formerly on-exchange listed Depository Receipts (DRs) is included. It is set to 1 if the DR continues to be traded OTC and 0 if it is withdrawn entirely from the U.S. stock market. The continued availability of DRs for OTC trading is expected to reduce the price reaction to the delisting announcement.

Lastly, two binary control variables are included to control for potential distortions associated with bear markets during the 2000-02 dotcom crisis and the 2008-09 financial crisis, i.e., the dummy variables CD and CF being 1 during these crisis periods and 0 otherwise. Table 1 provides a summary of the independent variables and their relationship to the hypotheses stated above.

---

Table 1: Variables used for hypothesis testing

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Variable</th>
<th>Variable (Abbreviation)</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td>Abnormal returns / cumulative abnormal returns</td>
<td>ARi/CARi</td>
<td>Abnormal return of listing i (AR, CAR_-3_1, CAR -3_3)</td>
</tr>
<tr>
<td>Market integration (H1)</td>
<td>Access of foreigners to local market</td>
<td>ACc</td>
<td>Access of foreigners to local market c (dummy variable equal to 1 if the EIU criteria is below 5, and 0 otherwise)</td>
</tr>
<tr>
<td>Repatriation</td>
<td>Repatriation of capital from country c</td>
<td>REPc</td>
<td>Repatriation of capital from country c (dummy variable equal to 0 if the EIU criteria is below 4, and 1 otherwise)</td>
</tr>
<tr>
<td>Correlation</td>
<td>Correlation of country c’s equity index with MSCI World index (measured over three year moving window of weekly returns)</td>
<td>CORRc</td>
<td>Correlation of country c’s equity index with MSCI World index (measured over three year moving window of weekly returns)</td>
</tr>
<tr>
<td>Maturity of local capital market (H2)</td>
<td>MVOL/GDP (local)</td>
<td>MVOLc</td>
<td>Market turnover of country c divided by GDP</td>
</tr>
<tr>
<td>Regulatory environment of local market (H3)</td>
<td>Investment protection</td>
<td>IPc</td>
<td>Investment protection schemes in country c (dummy variable equal to 0 if the EIU criteria is below 5, and 1 otherwise)</td>
</tr>
<tr>
<td>Transparency and fairness of legal system</td>
<td>Transparency and fairness of legal system (dummy variable equal to 0 if the EIU criteria is below 5, and 1 otherwise)</td>
<td>TFLSc</td>
<td>Transparency and fairness of legal system (dummy variable equal to 0 if the EIU criteria is below 5, and 1 otherwise)</td>
</tr>
<tr>
<td>Financial regulatory system</td>
<td>Financial regulatory system in country c</td>
<td>FRSc</td>
<td>Financial regulatory system in country c (dummy variable equal to 0 if the EIU criteria is below 5, and 1 otherwise)</td>
</tr>
<tr>
<td>Liquidity allocation (H4/H5)</td>
<td>Liquidity share in ADR</td>
<td>LSDRi</td>
<td>Share of total turnover generated in DR listing</td>
</tr>
<tr>
<td>Control variables</td>
<td>Total assets</td>
<td>TAi</td>
<td>Log of total assets of company i</td>
</tr>
<tr>
<td>Price-to-book ratio</td>
<td>Price to book ratio of company i</td>
<td>PTBi</td>
<td>Price to book ratio of company i</td>
</tr>
<tr>
<td>Return on assets</td>
<td>Return on assets of company i</td>
<td>ROAi</td>
<td>Return on assets of company i</td>
</tr>
<tr>
<td>Leverage ratio</td>
<td>Total leverage divided by total assets of company i</td>
<td>LEVRi</td>
<td>Total leverage divided by total assets of company i</td>
</tr>
<tr>
<td>Stddev (σ) (company)</td>
<td>Volatility of local capital market c (measured over three year moving window of weekly returns)</td>
<td>STDDEVc</td>
<td>Volatility of local capital market c (measured over three year moving window of weekly returns)</td>
</tr>
<tr>
<td>Post listing status (termination/OTC)</td>
<td>Dummy variable equal to 1 if company i’s shares are traded OTC after delisting and 0 otherwise</td>
<td>PLi</td>
<td>Dummy variable equal to 1 if company i’s shares are traded OTC after delisting and 0 otherwise</td>
</tr>
<tr>
<td>SEC-registration</td>
<td>Dummy variable equal to 1 if company i keeps its SEC registration and 0 otherwise</td>
<td>SECc</td>
<td>Dummy variable equal to 1 if company i keeps its SEC registration and 0 otherwise</td>
</tr>
<tr>
<td>Exchange (NYSE/NASDAQ)</td>
<td>Dummy variable equal to 1 if company i was listed on the NYSE and 0 otherwise</td>
<td>EXCHc</td>
<td>Dummy variable equal to 1 if company i was listed on the NYSE and 0 otherwise</td>
</tr>
<tr>
<td>Prior to rule 12h6</td>
<td>Dummy variable equal to 1 if the delisting took place prior to Rule12h-6 and 0 otherwise</td>
<td>R12h6i</td>
<td>Dummy variable equal to 1 if the delisting took place prior to Rule12h-6 and 0 otherwise</td>
</tr>
<tr>
<td>Voluntary/ involuntary</td>
<td>Dummy variables equal to 1 if the delisting is voluntary and 0 otherwise</td>
<td>VOLi</td>
<td>Dummy variables equal to 1 if the delisting is voluntary and 0 otherwise</td>
</tr>
<tr>
<td>Crisis (dotcom)</td>
<td>Dummy variables controlling for the dotcom crisis equal to 1 during the respective years and 0 otherwise</td>
<td>CD</td>
<td>Dummy variables controlling for the dotcom crisis equal to 1 during the respective years and 0 otherwise</td>
</tr>
<tr>
<td>Crisis (2007-08 financial)</td>
<td>Dummy variables controlling for the financial crisis equal to 1 during the respective years and 0 otherwise</td>
<td>CF</td>
<td>Dummy variables controlling for the financial crisis equal to 1 during the respective years and 0 otherwise</td>
</tr>
</tbody>
</table>

20 The inclusion of crises-related variables can be considered as an additional control. The concept of abnormal returns as presented above already implicitly controls for crises as it measures the stock price performance relative to a given market index which already captures the price-related effect of crises.
Accordingly, we specify the following regression model to explain the abnormal returns related to delistings:

\[ \text{AR} = \beta_0 + \sum_{i=1}^{k} \beta_i \text{Market}_i + \beta_0 \text{Maturity}_0 \text{Capital}_0 \text{Market} + \sum_{j=1}^{n} \beta_j \text{Environment}_j \text{Local}_0 \text{Capital}_0 \text{Market} + \sum_{z=1}^{2} \beta_z \text{Liquid}_z \text{Allocation} + \sum_{i=1}^{12} \beta_i \text{Control}_i \text{Variables} + \varepsilon_i \]  

(equation 8)

### 5. Empirical results

Tables 2 and 3 summarize the event study results based on weekly abnormal and cumulative abnormal returns of the two samples of companies locally listed in “developed” (Table 2) and in “emerging” markets (Table 3).

The statistical significance of abnormal and cumulative abnormal returns has been tested based on the four models previously defined, i.e., under the assumptions of cross-sectional independence and potential cross-sectional dependence, as well as by controlling for standardized abnormal returns and changes in variance related to the event itself.22

For the developed markets sample, all tests in Panel A of Table 2 indicate statistically significant negative abnormal returns during the announcement week as well as in the week following the announcement. Interestingly, the negative abnormal return of abnormal and cumulative abnormal returns has been tested on a per-stock level along the time dimension when computing the test statistic. The results of the non-parametric generalized sign test and the Corrado rank test are provided in sections B (weekly returns) and D (daily returns) of the Supplementary Internet Appendix available at http://www.sbf.unisg.ch/~/media/Internet/Content/Dateien/InstituteUndCenters/SBF/Mitarbeitende_CV_etc/Effects_from_Cross_Listings_Fuess_Hommel_Plagge_Internet_Appendix.ashx. Hereby, we again distinguish between companies locally listed in developed markets (Panel A) and those locally listed in emerging markets (Panel B). Two generalized sign tests as well as two Corrado rank tests have been applied. While the first set of tests is based on abnormal returns (Models I and III), the second set uses cumulative abnormal returns (Models II and IV).

---

### Table 2: Parametric event study results - developed markets

<table>
<thead>
<tr>
<th>Week</th>
<th>Model DM_I (AR) (cross-sectional independence)</th>
<th>Model DM_II (AR) (cross-sectional dependence)</th>
<th>Model DM_III (SAR)</th>
<th>Model DM_IV (SAR) (changes in variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha)</td>
<td>t-Statistic</td>
<td>(\alpha)</td>
<td>t-Statistic</td>
</tr>
<tr>
<td>-3</td>
<td>0.68% 0.0068 0.9996</td>
<td>0.68% 0.0059 1.1433</td>
<td>7.48% 0.0891 0.8394</td>
<td>7.48% 0.1103 0.6782</td>
</tr>
<tr>
<td>-2</td>
<td>1.07% 0.0068 1.5822</td>
<td>1.07% 0.0059 1.8096</td>
<td>13.96% 0.0891 1.5666</td>
<td>13.96% 0.1166 1.1967</td>
</tr>
<tr>
<td>-1</td>
<td>-0.49% 0.0068 -0.7313</td>
<td>-0.49% 0.0059 -0.8364</td>
<td>-7.81% 0.0891 -0.8771</td>
<td>-7.81% 0.0968 -1.1188</td>
</tr>
<tr>
<td>0</td>
<td>-1.18% 0.0068 -1.7489</td>
<td>-1.18% 0.0059 -2.0002</td>
<td>-16.35% 0.0891 -1.8356</td>
<td>-16.35% 0.1066 -1.6341</td>
</tr>
<tr>
<td>1</td>
<td>-2.11% 0.0068 -3.1173</td>
<td>-2.11% 0.0059 -3.5653</td>
<td>-25.14% 0.0894 -2.8110</td>
<td>-25.14% 0.1049 -2.3976</td>
</tr>
<tr>
<td>2</td>
<td>0.37% 0.0068 0.5526</td>
<td>0.37% 0.0059 0.6320</td>
<td>-6.93% 0.0894 -0.7749</td>
<td>-6.93% 0.1239 -0.5592</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Week</th>
<th>Model DM_V (CAR) (cross-sectional independence)</th>
<th>Model DM_VI (CAR) (cross-sectional dependence)</th>
<th>Model DM_VII (SCAR)</th>
<th>Model DM_VIII (SCAR) (changes in variance)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha)</td>
<td>t-Statistic</td>
<td>(\alpha)</td>
<td>t-Statistic</td>
</tr>
<tr>
<td>-3</td>
<td>0.68% 0.0068 0.9996</td>
<td>0.68% 0.0059 1.1433</td>
<td>6.80% 0.0891 0.7632</td>
<td>6.80% 0.1072 0.6340</td>
</tr>
<tr>
<td>-2</td>
<td>1.74% 0.0096 1.8257</td>
<td>1.74% 0.0084 2.0880</td>
<td>14.62% 0.1260 1.1604</td>
<td>14.62% 0.1540 0.9490</td>
</tr>
<tr>
<td>-1</td>
<td>1.25% 0.0117 1.0684</td>
<td>1.25% 0.0102 1.2220</td>
<td>7.47% 0.1543 0.4840</td>
<td>7.47% 0.1831 0.4080</td>
</tr>
<tr>
<td>0</td>
<td>0.07% 0.0135 0.0509</td>
<td>0.07% 0.0118 0.0582</td>
<td>-1.78% 0.1782 -0.0999</td>
<td>-1.78% 0.2088 -0.0853</td>
</tr>
<tr>
<td>1</td>
<td>-2.02% 0.0151 -1.3375</td>
<td>-2.02% 0.0132 -1.5298</td>
<td>-12.58% 0.1992 -0.6313</td>
<td>-12.58% 0.2331 -0.5396</td>
</tr>
<tr>
<td>2</td>
<td>-1.65% 0.0165 -0.9972</td>
<td>-1.65% 0.0145 -1.1405</td>
<td>-14.29% 0.2182 -0.6547</td>
<td>-14.29% 0.2551 -0.5600</td>
</tr>
</tbody>
</table>
during the event week is less significant when taking into account changes in variance related to the event itself (Model IV). This can be interpreted as a first indication of an increase in variance due to the delisting event.

As shown in Panel B of Table 2, only two of the four parametric models show cumulative abnormal returns that are significantly negative in the week following the announcement. When considering changes in variance, no statistically significant results could be observed. This result for the event week can be explained by the positive abnormal returns in the event window prior to the event week. Even though there is a sharp decrease in abnormal returns starting with the event week itself, it takes an additional week to mitigate the effect of the positive abnormal returns in the pre-delisting period when considering cumulative abnormal returns.

Even though the t-statistics of the generalized sign tests provide no significant results, the proportion of negative abnormal returns in a given event week (p-values of Model DM_I) increase from 52.38 to 59.2 over the range of six weeks around the delisting event. They reach their peak in the pre-announcement week with a share of 60.32% of negative abnormal returns. Similar results can be found for the respective proportion of cumulative negative abnormal returns (model DM_II). Unlike the generalized sign test, the Corrado (1989) rank test also takes into account the magnitude and not merely the sign of the abnormal returns. It documents statistically significant negative abnormal and cumulative abnormal returns as of the first week following the announcement week (Models DM_III and DM_IV). The fact that the rank test provides significant results, whereas the sign test fails to do so, indicates that the negative abnormal returns observed by the parametric tests are most likely caused by the magnitude of a few negative observations than by a consistent decrease in the stock prices across all stocks in the sample.

The observation of negative abnormal returns based on the developed market sample is surprising, as it was assumed that these companies already have good access to the international capital market via their local exchange. In addition, they should be subjected to reasonably high local disclosure standards that make bonding less significant. Consequently, it could be argued that the cost reduction associated with the cross-delisting should have led to positive abnormal returns. Summing up at this stage of the analysis, it can already be stated that even for companies listed in developed markets, a delisting viewed negatively by the market.

For the emerging market sample of Table 3, we find significant negative abnormal returns in the announcement week only in the models for standardized abnormal returns (Models EM_III and EM_IV). However, a graphical representation of the abnormal and cumulative abnormal returns based on the two samples (developed and emerging markets) reveals interesting insights.

Panel A of Figure 1 shows that stocks locally listed in emerging markets experience on average significantly negative abnormal returns two weeks prior to the week of the delisting event. Consequently, their cumulative abnormal returns for the entire event period, as displayed in Panel B of Figure 1, are much lower than the developed market sample. The reason why the cumulative abnormal returns are not consistently significant for the respective samples can be explained by inspecting the development of the t-statistics over time as displayed in Figure 2.

From the graphical representation, it can be concluded that the standardization of abnormal returns tends to generally decrease the statistical significance of cumulative abnormal returns. Taking into account changes in variance related to the event period further increases the standard deviation used to calculate the test statistic, which in turn additionally decreases statistical significance. The fact that the t-values of both samples (developed and emerging markets) at the end of the event period are on comparable levels, even though the cumulative abnormal returns for the emerging market sample are significantly lower, can be explained by much higher volatility levels within the emerging market, as compared to the developed market sample (see standard deviations for cumulative abnormal returns as reported in Tables 2 and 3). Additionally, when testing the significance of cumulative abnormal returns, the standard deviation grows with the square root of time which, ceteris paribus, further decreases statistical significance over time.

23 See Panel A of Section B of the Supplementary Internet Appendix (available at http://www.sbf.unisg.ch/~media/Internet/Content/Dateien/Institute/indContent/SBF/Mitarbeitende_CV/etc/Effects_from_Cross_Listings_Fuess_Hommel_Flagge_Internet_Appendix.ashx).
Similar results are found for the analyses based on daily abnormal returns. With the exception of the model that takes into account event-related changes in variance, all models display negative abnormal returns on the announcement day for the developed market sample. Further, significantly negative returns can be observed during the days following the delisting announcement. As a next step, we analyze the driving factors of stock price movements related to cross-delistings by using the estimated weekly abnormal returns as dependent variables in the regression models of Table 4. In order to test the robustness of the empirical results, the regressions are run with and without the previously defined control variables.

As a next step, we analyze the driving factors of stock price movements related to cross-delistings by using the estimated weekly abnormal returns as dependent variables in the regression models of Table 4. In order to test the robustness of the empirical results, the regressions are run with and without the previously defined control variables. To test their robustness, the regressions are repeated based on three different models.

The following analysis focuses on the extended Models Ib, Ib and IIIb, including control variables. In case the results between both models, parsimonious and extended, significantly deviate, it is explicitly mentioned at the end of the analysis.

26 See graphical representations provided in Section C of the Supplementary Internet Appendix available at http://www.sbf.unisg.ch/~media/Internet/Contents/Dateien/InstituteUndCenters/SBF/Mitarbeitende_CV_etc/ Effects_from_Cross_Listings_Fuess_Hommel_Plagge_Internet_Appendix.ashx.

27 The following analysis focuses on the extended Models Ib, Ib and IIIb, including control variables. In case the results between both models, parsimonious and extended, significantly deviate, it is explicitly mentioned at the end of the analysis.

See Sections C and D of the Supplementary Internet Appendix available at http://www.sbf.unisg.ch/~media/Internet/Contents/Dateien/InstituteUndCenters/SBF/Mitarbeitende_CV_etc/ Effects_from_Cross_Listings_Fuess_Hommel_Plagge_Internet_Appendix.ashx.

24 E.g., the standard deviation of the model assuming cross-sectional independence is higher by a factor 2.37 than the standard deviation of the developed market sample $\sigma_{(AR,D)} = 0.0064$ (EM) versus $\sigma_{(AR,D)} = 0.0027$ (DM).
dependent variables covering different time periods around the announcement date. The first and second columns include the regression coefficients for cumulative abnormal returns based on a six-week period surrounding the event date (Model I). Columns three and four report the results for cumulative abnormal returns based on a four-week period starting 3.5 weeks prior to the event week and ending with the event week (Model II). Finally, the last two columns provide the regression coefficients for abnormal returns measured in the event week itself (Model III).

Table 5 further provides the correlation coefficients of the independent variables. The coefficients range from a minimum of -0.66 between the standard deviation (STDDEV) and the return on assets (RoA) and a maximum of 0.61 between the variables investment protection (IP) and market correlation (CORR).28

Market integration: in support of H1, Model III identifies the access of foreigners as a relevant parameter to determine the magnitude of abnormal returns in the announcement week. Abnormal returns for companies domiciled in freely accessible

---

28 The negative correlation between STDDEV and RoA indicates that highly volatile companies tend to be less profitable. The positive correlation between IP and CORR makes intuitive sense. Integrated countries have higher investment protection as compared to those with underdeveloped property rights. To control for multicolinearity, variance inflation factors have been calculated for each independent variable. However, no factor was found to exceed a value of 4.5, implying that the model should not suffer from multicollinearity issues.
Table 4: Regression results

This table provides the regression coefficients and robust standard errors for equation (8) based on the core models (i.e., without control variables) as well as the full specification (i.e., including controls). Models Ia and Ib include the results of the regression based on cumulative abnormal returns (CAR) for a six-week period ranging from d_0–3.5 weeks to d_0+2.5 weeks, Models IIA and IIB display the regression results for cumulative abnormal returns (CAR) ranging from d_0–0.5 weeks to d_0+0.5 weeks, and Models IIIa and IIIb provides the regression results for abnormal return (AR) in the event week (d_0–0.5 weeks to d_0+0.5 weeks). All values are multiplied by 100 for better reading. a, b and c indicate statistical significance at the 1%, 5% and 10% level, respectively.

<table>
<thead>
<tr>
<th>Hypothesis/Control variables</th>
<th>Variable</th>
<th>Model Ia (CAR: d_0–3.5 weeks to d_0+2.5 weeks)</th>
<th>Model Ib (CAR: d_0–3.5 weeks to d_0+0.5 weeks)</th>
<th>Model IIA (CAR: d_0–0.5 weeks to d_0+0.5 weeks)</th>
<th>Model IIB (AR: d_0–0.5 weeks to d_0+0.5 weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market Integration (H1)</td>
<td>AC_{t-1}</td>
<td>3.8153 (3.6062)</td>
<td>1.2330 (3.667)</td>
<td>2.0093 (2.8330)</td>
<td>-1.0787 (2.9494)</td>
</tr>
<tr>
<td></td>
<td>REP_{t-1}</td>
<td>-2.4719 (5.3859)</td>
<td>3.9041 (5.9233)</td>
<td>1.9513 (5.2162)</td>
<td>7.0011 (5.5944)</td>
</tr>
<tr>
<td></td>
<td>CORR_{t-1}</td>
<td>-1.1239 (14.1999)</td>
<td>-7.4430 (12.4765)</td>
<td>-2.1659 (12.8676)</td>
<td>-14.2561 (13.1203)</td>
</tr>
<tr>
<td>Maturity of local capital market (H2)</td>
<td>MVOL_{t-1}</td>
<td>0.0022 (0.0170)</td>
<td>0.0042 (0.0129)</td>
<td>0.0017 (0.0128)</td>
<td>0.0012 (0.0101)</td>
</tr>
<tr>
<td>Regulatory environment of local market (H3)</td>
<td>IP_{t-1}</td>
<td>2.21278 (2.84966)</td>
<td>3.8293 (3.7566)</td>
<td>0.0124 (2.8315)</td>
<td>2.42236 (3.2509)</td>
</tr>
<tr>
<td></td>
<td>TFLS_{t-1}</td>
<td>2.75856 (3.00298)</td>
<td>3.5796 (4.2386)</td>
<td>0.6000 (2.2897)</td>
<td>1.2361 (3.1524)</td>
</tr>
<tr>
<td></td>
<td>FRS_{t-1}</td>
<td>-12.06863b (5.3882)</td>
<td>-12.7649c (6.7380)</td>
<td>-10.0755 (5.1332)</td>
<td>-7.8159 (5.7260)</td>
</tr>
<tr>
<td>Liquidity allocation (H4/H5)</td>
<td>LSDR_{t-1}</td>
<td>-31.7739 (8.6835)</td>
<td>-32.5847a (8.4197)</td>
<td>-16.7230 (9.1046)</td>
<td>-22.1177b (10.0382)</td>
</tr>
<tr>
<td></td>
<td>LSEM_{t-1}</td>
<td>2.9394 (8.300)</td>
<td>6.9182 (8.2184)</td>
<td>-0.0839 (8.2538)</td>
<td>2.7239 (7.6541)</td>
</tr>
<tr>
<td>Control variables</td>
<td>ln(TA)_{t-1}</td>
<td>0.0515 (0.8409)</td>
<td>0.1141 (0.7008)</td>
<td>0.1141 (0.7008)</td>
<td>0.0789 (0.3579)</td>
</tr>
<tr>
<td></td>
<td>PTB_{t-1}</td>
<td>-0.5511 (0.3427)</td>
<td>-0.2856 (0.3119)</td>
<td>-0.2856 (0.3119)</td>
<td>-0.0381 (0.1194)</td>
</tr>
<tr>
<td></td>
<td>ROA_{t-1}</td>
<td>-0.2630a (0.1124)</td>
<td>-0.1540 (0.1061)</td>
<td>-0.1540 (0.1061)</td>
<td>0.1561 (0.1075)</td>
</tr>
<tr>
<td></td>
<td>LEVR_{t-1}</td>
<td>-3.2346 (6.4818)</td>
<td>-8.8130 (4.6203)</td>
<td>-8.8130 (4.6203)</td>
<td>-2.1859 (3.8988)</td>
</tr>
<tr>
<td></td>
<td>STDDEV_{t-1}</td>
<td>-50.0101 (40.7723)</td>
<td>-11.9929 (29.9215)</td>
<td>-11.9929 (29.9215)</td>
<td>56.3752b (21.6338)</td>
</tr>
<tr>
<td></td>
<td>PL_{t-1}</td>
<td>-4.4880 (4.7983)</td>
<td>-3.6607 (4.5293)</td>
<td>-3.6607 (4.5293)</td>
<td>-2.5280 (2.6288)</td>
</tr>
<tr>
<td></td>
<td>SEC_{t-1}</td>
<td>-10.3245a (4.8769)</td>
<td>-6.6407 (3.7136)</td>
<td>-6.6407 (3.7136)</td>
<td>-1.9796 (1.6861)</td>
</tr>
<tr>
<td></td>
<td>EXCH_{t-1}</td>
<td>-1.1690 (3.3642)</td>
<td>-0.8679 (3.3182)</td>
<td>-0.8679 (3.3182)</td>
<td>-1.6403 (1.7609)</td>
</tr>
<tr>
<td></td>
<td>R12h6_{t-1}</td>
<td>1.1739 (5.2036)</td>
<td>4.9209 (5.0282)</td>
<td>4.9209 (5.0282)</td>
<td>-0.1013 (2.6507)</td>
</tr>
<tr>
<td></td>
<td>VOL_{t-1}</td>
<td>-0.9043 (4.1453)</td>
<td>4.1281 (4.2425)</td>
<td>4.1281 (4.2425)</td>
<td>4.9275b (2.6350)</td>
</tr>
<tr>
<td></td>
<td>CD_{t-1}</td>
<td>2.2125 (5.4733)</td>
<td>2.3342 (5.3837)</td>
<td>2.3342 (5.3837)</td>
<td>-1.5205 (3.5918)</td>
</tr>
<tr>
<td></td>
<td>CF_{t-1}</td>
<td>9.5762 (9.3416)</td>
<td>7.3294 (6.8419)</td>
<td>7.3294 (6.8419)</td>
<td>1.1413 (1.9831)</td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>-0.7902 (8.7782)</td>
<td>-2.9494 (18.1522)</td>
<td>-0.4672 (8.4841)</td>
<td>-1.9102 (14.9979)</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td></td>
<td>0.1164 (0.2020)</td>
<td>0.0723 (0.1867)</td>
<td>0.0723 (0.1867)</td>
<td>0.1031 (0.237)</td>
</tr>
</tbody>
</table>
countries are on average 2.89 percentage points higher than for those listed in countries difficult to access by foreign investors. It is reflective of the fact that the return to a segmented or semi-integrated market leads to a cost of capital increase. The variables correlation (CORR) and capital repatriation (REP) do not exert a significant influence on abnormal returns. Similarly, the market integration proxies are not found to affect cumulative abnormal returns (Models I and II). The lack of significance might be explained by a distortion of cumulative abnormal returns resulting from price reactions prior to the delisting week as observed in Figure 2 in the context of the previous discussion.

Capital market liquidity: the trading volume scaled by GDP, as a proxy for the maturity of the local capital market, is found to positively influence abnormal stock price movements in the delisting week. An increase in the aggregated turnover of the local market by one standard deviation increases the abnormal return by 0.99 percentage points. As hypothesized in H2, companies that are locally listed in mature capital markets suffer less from a foreign delisting. Again, the variables’ influence on cumulative abnormal returns (Models I and II) is insignificant.

Regulatory environment of the local market: among the variables used as proxies for the regulatory environment, only the transparency and fairness of the legal system (TFLS in Model III) and the financial regulatory system (FRS in Model II) show a statistically significant influence on abnormal returns around the delisting announcement. Interestingly, their influence contradicts H3. Companies locally listed in countries that have a more transparent and fairer legal system (better financial regulatory system), on average, experience lower abnormal returns than those locally listed in markets with a lower TFLS (FRS) score. One explanation for this negative relationship might be that companies resident in markets with high regulatory standards experience lower delisting-related savings as they need to maintain high standards of disclosure requirements.

Table 5: Correlation matrix of independent regression variables

Correlation coefficients of all explanatory variables used in the regression as defined in equation (8). The coefficients of binary variables are calculated according to Spearman rank correlation coefficient.
Liquidity allocation: in support of H4, the share of total liquidity generated in the DR is found to have a significant and negative influence on abnormal and cumulative abnormal stock price movements. This observation is consistent across all three models. An increase in the share of DR liquidity by one standard deviation (18.09%) decreases abnormal returns by 2.42 percentage points and cumulative abnormal returns by as much as 5.90 percentage points (e.g., as measured in Model I). Thus, a higher relative DR liquidity, as an indicator for the importance of a foreign investor base of a given company, leads to lower abnormal returns. As a consequence of the foreign delisting, such companies suffer from higher negative abnormal returns after returning to their local investor base. Other than assumed in H5, the fraction of total liquidity generated in emerging markets prior to the delisting is not found to have a significant influence on delisting-related stock price movements. The results rather indicate that, after controlling for the accessibility of the local capital market, the liquidity attached to the DR listing itself is of value irrespective of the overall liquidity spread between developed and emerging market places.

Control variables: among the control variables included in the model only a few are found to be significant. As expected, the leverage ratio is negatively related to abnormal returns (Models II). An increase in the companies’ indebtedness of one standard deviation (26.04%) leads to a decrease in cumulative abnormal returns of 2.30 percentage points (Model II). Thus, the market negatively values the delisting attempt for companies that are relatively more indebted. In case of highly indebted companies, a foreign delisting might even signal financial distress to outside investors, which in turn leads to relatively higher negative abnormal returns. However, this conjecture is only supported by the second model, whereas the influence of leverage is not significant in Models I and III.

Companies’ total risk, as measured by their stock price volatility, is found to exert a positive influence on abnormal stock price movements (Model III). This observation is in line with the initial assumption indicating that the amplitudes caused by a delisting announcement might be more pronounced for highly volatile stocks. RoA influences cumulative abnormal returns (Model I) negatively. Thus, the more profitable a company is, the less positive is the market’s reaction to the delisting announcement. In light of the loss of competitiveness argument stated above, this observation makes intuitive sense. The decrease in administrative costs associated with a delisting is likely to relax financing constraints for less profitable companies. In addition, companies announcing that they intend to maintain their SEC registration experience lower abnormal returns than to those choosing to deregister. In the face of the competitiveness argument, this indicates that the market values cost savings more favorably than the continuation of the SEC registration. Lastly, companies that voluntarily delist show less negative abnormal returns than those that involuntarily delist. A self-controlled action, presumably taken after a careful evaluation of associated risks, should be more positively valued by investors than a delisting enforced by a third party.29

6. Conclusion
The event study results show that the market reaction to a delisting announcement is, on average, negative across all companies. Over the course of the different specifications of event periods, the negative price reactions are even found to accumulate. On the event date itself, this reaction is statistically significant for the developed market sample. However, for the emerging market sample, it is not found to be consistently significant across all statistical tests applied, even though the cumulative abnormal returns are more negative compared to the developed market sample. This missing significance can mainly be attributed to the significantly higher volatility of abnormal returns within the emerging market sample, which leads to biased test statistics when not adjusted for increasing variance during the event window.

The observation of consistent negative abnormal returns related to the announcement of cross-delistings shows that a foreign listing is generally viewed positively by market participants. This evaluation even holds for companies that are locally listed on already well-integrated stock exchanges. However, companies locally listed in emerging markets suffer, on average, higher negative abnormal returns. This empirical observation is

29 The parsimonious model underscores the significance of investors’ access to the local market (AF) when evaluating abnormal returns (Model III). The magnitude of its influence even increases compared to the extended model. The negative influence of FRS and TFLS on abnormal and cumulative abnormal returns is supported as well. Lastly, the comparison of the two models underscores the significance of the share of total liquidity concentrated in the DR prior to the delisting announcement. The higher the share of DR liquidity (LSDR), the more negative is the observed price reaction.
consistent with the hypothesis that companies which cannot rely on an international investor base but are mainly restricted to local investors are faced with higher refinancing costs. Local investors of segmented or semi-integrated markets cannot diversify away the local country's specific risk entirely and consequently require a premium for carrying additional risk. Thus, companies that return to segmented or semi-integrated local capital markets experience a relatively higher devaluation of their stocks in anticipation of higher refinancing costs in the future.

These results are further supported by the regression analysis to explain abnormal returns. The ease of access to a given market is found to exert a positive influence on abnormal returns associated with the delisting announcement. Furthermore, the share of total liquidity allocated to the foreign listing is negatively related to abnormal performance. If a company withdraws its cross-listing that generates a high portion of its liquidity in a foreign listing, it loses an essential access route to a given market, and that is viewed negatively by its investors.

References
About EY
EY is a global leader in assurance, tax, transaction and advisory services. The insights and quality services we deliver help build trust and confidence in the capital markets and in economies the world over. We develop outstanding leaders who team to deliver on our promises to all of our stakeholders. In so doing, we play a critical role in building a better working world for our people, for our clients and for our communities.

EY refers to the global organization, and may refer to one or more, of the member firms of Ernst & Young Global Limited, each of which is a separate legal entity. Ernst & Young Global Limited, a UK company limited by guarantee, does not provide services to clients. For more information about our organization, please visit ey.com.

© 2014 EYGM Limited.
All Rights Reserved.

EYG No. CQ0106

In line with EY’s commitment to minimize its impact on the environment, this document has been printed on paper with a high recycled content.

This material has been prepared for general informational purposes only and is not intended to be relied upon as accounting, tax, or other professional advice. Please refer to your advisors for specific advice.

ey.com

The articles, information and reports (the articles) contained within The Journal are generic and represent the views and opinions of their authors. The articles produced by authors external to EY do not necessarily represent the views or opinions of EYGM Limited nor any other member of the global EY organization. The articles produced by EY contain general commentary and do not contain tailored specific advice and should not be regarded as comprehensive or sufficient for making decisions, nor should be used in place of professional advice. Accordingly, neither EYGM Limited nor any other member of the global EY organization accepts responsibility for loss arising from any action taken or not taken by those receiving The Journal.

Accredited by the American Economic Association

ISSN 2049-8640